October 2018

Estimation in High-Dimensional Factor Models with Structural Instabilities

Wen Gao
University of Windsor, gao135@uwindsor.ca

Estimation in High-Dimensional Factor Models with Structural Instabilities

by

Wen (Wendy) Gao

A Major Research Paper
Submitted to the Faculty of Graduate Studies
through the Department of Mathematics and Statistics
in Partial Fulfillment of the Requirements for
the Degree of Master of Science at the
University of Windsor

Windsor, Ontario, Canada

Estimation in High-Dimensional Factor Models with Structural Instabilities

by

Wen (Wendy) Gao

APPROVED BY:

A. Hussein
Department of Mathematics and Statistics

S. Nkurunziza, Advisor
Department of Mathematics and Statistics

September 18, 2018

Author's Declaration of Originality

I hereby certify that I am the sole author of this major paper and that no part of this major paper has been published or submitted for publication.

I certify that, to the best of my knowledge, my major paper does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my major paper, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my major paper and have included copies of such copyright clearances to my appendix.

I declare that this is a true copy of my major paper, including any final revisions, as approved by my major paper committee and the Graduate Studies office, and that this major paper has not been submitted for a higher degree to any other University or Institution.

Abstract

In this major paper, we use high-dimensional models to analyze macroeconomic data which is influenced by the break point. In particular, we consider to detect the break point and study the changes of the number of factors and the factor loadings with the structural instability.

Concretely, we propose two factor models which explain the processes of pre- and post- break periods. Then, we consider the break point as known or unknown. In both situations, we derive the shrinkage estimators by minimizing the penalized least square function and calculate the estimators of the numbers of pre- and post- break factors and the existence of the break point. After that, we present some results about the asymptotic performance of the penalty least square estimators with both the cross-section dimension and the time dimension tend to infinity.

In addition, we establish the Monte Carlo simulation to evaluate the performance of the procedure and we analyze the real dataset from 2007-2009 Great Recession. The theoretical results are confirmed by simulation that the break point can be properly detected. More than half proposed post model selection estimators domain the full sample estimators while the procedure performs relatively poor in estimating the number of pre- and post- factors.

To my loving parents

Yun Gao and Bingmin Lian

Acknowledgments

First I would like to express my sincere gratitude to my supervisor Dr. Sévérien Nkurunziza. The knowledge I learned from him and the continuous support he gave made me finally succeed in this work. His devotion toward academic research and rigorous standard of major paper motivated me and gave me confidence to pursue higher-quality work. During my program, he was like a leader of my academic road, inspiring my enthusiasm to dig deeply and get a better understanding of Statistics.

I would also like to thank Dr. Abdulkadir Hussein for being my department reader. Without their useful advice, this major paper cannot be completed and improved. His knowledge of data analysis helped me build a solid foundation for further studies in Statistics.

In addition, I would like to thank all faculty and staff members as well as the graduate students in the Department of Mathematics and Statistics who provided me assistant during my study.

Moreover, my profound appreciation goes to my parents. They have given me constant love, encouragement and support since my birth. Most importantly, they have guided me to face my life positively and have always been the solid backing for me. Thank you!

Contents

List of Tables

Chapter 1

Introduction

1.1 Background

This major paper is about estimation and modeling of macroeconomic data in a high-dimensional setting. In macroeconomics, it is common to study the economic system and its influence on the financial market in a long-term. In particular, the aim of studying macroeconomics is to understand the reason for the economic fluctuations, to achieve persistent economic growth and to maximize the level of employment and Gross Domestic Product (GDP) in a country (Giannone et al. 2008)[20].

The macroeconomics phenomena that we studied in this major paper is under the high-dimensional setting. Briefly, high-dimensional data refers to the situation where the number of unknown parameters is much larger than the sample size. Such a model is common in different areas of macroeconomics (Fan et al. 2011)[17]. Vector autoregressive (VAR) model (Sims et al. 1990)[29] is one example of the high dimensional models, which is considered to be an effective method to evaluate the joint evolution of macroeconomic time series. However, macroeconomic problems are usually associated with hundreds of data series, which may result in quadratic or cubic growth in the number of parameters. To reduce the difficulty of estimating large number of parameters, Bernanke (2005)[7] proposed an augment standard VARs with estimated factors (FAVAR) to evaluate the consequences of monetary policy (Fan et al. 2011)[17]. Besides that, another difficulty of analyzing high dimensional data is that we have to cope with measures of uncertainty or stability (B$\ddot{u}$hlmann and Van de Geer, 2011)[11]. The additional difficulty we focus on in this major paper is the potential break of the economic or financial period. Indeed, the analysis of macroeconomics can be interrupted by structural breaks such as the 2007-2009 Great Recession. However, the structural breaks is generally unknown, which makes the statistical analysis more complicated.

In terms of a real-world application, many intriguing questions emerge: whether the Great Recession or other post-war U.S. recessions cause a change in the business cycle or not, are these breaks always accompanied by the emergence of new factors such as financial or credit factors, or are there any changes to the existing factors? (Cheng et al. 2016)[14]. This major paper is generated to answer these questions. It locates the break point if it is unknown, captures any new factors and estimates the change of the factors after the break point.

1.2 Existing Studies

Throughout the research, we reviewed the existing papers which use different techniques to analyze the problems mentioned above. The methods include locating the break point, evaluating changes of factor loadings and estimating the number of factors before and after the break point. For example, Bai (1997)[4] uses conventional residual-based procedures to estimate the break point, but his method is based on the known number of factors. Other studies given in Bai and Ng (2002)[5], Onatski (2010)[26], or Ahn and Horenstein (2013)[1] require the break point to be known and they are not able to capture the changes in factor loadings. Other works are given in Breitung and Eickmeier (2011)[10], Chen et al. (2014)[13], or Han and Inoue (2014)[22]. They all neglect the estimation of the number of factors as well as the emergence of the new factors in the process of structural break tests. Moreover, the methods in Stock and Watson (2012)[31] cannot work without the knowledge of the break point.

1.3 Improvements

Cheng et al.(2016)[14] improved over the references mentioned above by dealing with the situation that the break point is unknown. The proposed method simultaneously estimates the break point if it is unknown and determines the numbers of pre- and post-break factors. By the improved techniques, the overestimation of the number of factors is avoided. In this case, a factor model instability is considered. The definition of the instability refers to the large changes in the factor loadings when the number of factors is constant.

The method used in this major paper applies two identification results for the factor model. First, this major paper shows that if the space spanned by the factor loadings or the scaling of the factor loadings changes, then a structural change is identifiable. Second, if the break point is unknown, then it could be determined by the factor model. Then, the number of pre- and post-break factors can be estimated more precisely under the knowledge of the break point.

The penalized least squares (PLS) method is used to obtain the estimator. The group-Least Absolute Shrinkage and Selection Operator (LASSO) penalties are applied to pre-break factor loadings and the change in loadings. A PLS criterion function is developed, and the estimator is generated from the minimization of this function.

1.4 Organization of the Major Paper

The remainder of this major paper is organized as follows. Chapter 2 introduces two versions of the factor models with the structural instability characterized by the knowledge of the break point. Chapter 3 demonstrates the whole process of deriving and calculating the shrinkage estimators, the mathematical algorithm applied to improve our work and the asymptotic result for the large sample behavior of the factors under the known break point case. The situation when the break point is unknown is explained in Chapter 4. Chapter 5 is separated into two parts. Part one displays the design and the result of the Monte Carlo Simulations about the performance of our estimators. Part two of Chapter 5 implements our procedure to real world macroeconomic problem, the Great Recession. Finally, the result of our findings are concluded in Chapter 6. The proof of one major theorem is attached in the Appendix.

Chapter 2

The Statistical Model

In this Chapter, Section 2.1 gives an overview about the two versions of the factor models which are designed for pre- and post-break periods respectively. Section 2.2 briefly explains the identification of a structural change under known or unknown cases.

Suppose that $t \in \{1, 2, \cdots, T_0, \cdots, T\}$, which denotes a set of T time points. T_0 is the only break point which is usually unknown. Suppose that there are r_a unobserved pre-break factors at T_0, and r_b unobserved post-break factors after T_0. We assume we have the data $\{X_{it} \in R : i = 1, \cdots, N, t = 1, \cdots, T\}$ for the i-th cross-section unit at time t. Note that N and T are the cross-section dimension and the time dimension of the data set respectively (Bai and Ng, 2002)[5]. $X_t = (X_{1t}, \cdots, X_{Nt})' \in R^{N\times 1}$ are the observations at time t. Two different versions of factor model are generated for the periods before and after the break point which will be discussed in the following Section.

2.1 The factor Model

This section introduces the formation of the factor models with the observations X, the unknown factor loadings and the unknown parameters. The models have two different versions for the data observed before or after the break point.

The whole process is split into two parts by the potential structural break point. Let $F_a = (F_1^0, \cdots, F_{T_0}^0)' \in R^{T_0 \times r_a}$ denote the factors in the pre-break period, and let Λ^0 stand for an $N \times r_a$ matrix which denotes the factor loadings associated with F_a. For the time points at or before T_0, the model of pre-break period is

$$X_a = F_a \Lambda^{0\prime} + e_a, \tag{2.1}$$

where $X_a = (X_1, \cdots, X_{T_0})' \in R^{T_0 \times N}$, and $e_a = (e_1, \cdots, e_{T_0})' \in R^{T_0 \times N}$, which is the error term. The factors, their loadings and the errors are not observable.

For the post-break model, the matrix F_b are separated into $F_{b,1}$ and $F_{b,2}$, i.e., $F_b = (F_{b,1}, F_{b,2})$. $F_{b,1} \in R^{T_1 \times r_a}$ stands for the pre-break factors in the post-break period, and $F_{b,2} \in R^{T_1 \times (r_b - r_a)}$ stands for the new factors that emerge after the break point. Let $\Gamma^0 = (\Gamma_1^0, \Gamma_2^0)$ denote the change in the factor loadings, let Γ_1^0 denote the change of the pre-break factors F_t^0 and let Γ_2^0 denote the change of the new factors F_t^*. The matrix form of post-break model is given as follows

$$X_b = F_{b,1}(\Lambda^0 + \Gamma_1^0)' + F_{b,2}\Gamma_2^{0\prime} + e_b, \tag{2.2}$$

where $X_b = (X_{T_0+1}, \cdots, X_T)'$, $F_{b,1} = (F^0_{T0+1}, \cdots, F^0_{T_0})'$, $F_{b,2} = (F^*_{T_0+1}, \cdots, F^*_T)'$ and $e_b = (e_{T_0+1}, \cdots, e_T)'$, which is the error term.

2.2 Instability

In this section, we introduce the structural instability. In this major paper, we assume that the number of post-break factors is always greater than or equal to pre-break factors, which means $r_b \geq r_a$.

The structural instability is defined as following:

$$r_a = r_b \text{ and } \Gamma^0_1 \neq 0. \tag{2.3}$$

Under the instability, the number of factors remains the same after the break point, but the factor loadings changes.

Chapter 3

Preliminary Results: Known Break Point Case

In the model we discussed in the previous chapter, there are four critical parts, the structural instability, the break point, the change of factor loadings, and the number of factors. Above all, we analyze the instability, the change of factor loadings, and the number of factors given that the break point is known. First, we show the identification of the instability and then introduce the shrinkage estimator. Next, we show how we obtain the estimators of the occurrence of break point and the number of factors, which are determined by the shrinkage estimator. Then, we apply a two-step estimation algorithm to analyze the shrinkage estimator with adjusted penalty weights done by cross validation. Finally, we present the asymptotic theory for the large sample behavior of the factors.

In the first Section, we briefly explain the way to identify the instability. In Section

3.2, we derive the shrinkage estimator and highlight its importance in the process of this analysis. Section 3.3 introduces the estimators which contains the estimators of break point and the number of factors. The technique of deriving the post model selection estimator, which can optimize the penalty terms of the shrinkage estimator, is demonstrated in Section 3.4. Section 3.5, 3.6 and 3.7 present the algorithm to calculate the shrinkage estimator. The assumptions and theorems associated with the large sample are clarified in Section 3.8. In practice, the break point is usually unknown, and in this case, the statistical analysis is more complex. We will generalize the results from known break point to unknown break point in the next chapter.

3.1 Identification of the Instability

Since the break point is known, the number of pre- and post-break factors r_a and r_b can be analyzed using the model selection method constructed by Bai and Ng (2002)[5]. To identify the instability, first, we need to have the knowledge of the normalization of the factor covariance matrix. Second, we show that either the space spanned by the factor loadings or the scaling of the factor loadings changes (Cheng et al. 2016)[14]. We define an augmented covariance matrix with dimension $(r_a + r_b) \times (r_a + r_b)$:

$$\Sigma_{\Lambda\Psi}^{+} = \begin{bmatrix} \Sigma_{\Lambda} & \Sigma_{\Lambda\Psi} \\ \Sigma'_{\Lambda\Psi} & \Sigma_{\Psi} \end{bmatrix}. \tag{3.1}$$

We can identify the structural instability by using the following Assumption 1 from Cheng et al. (2016)[14].

Assumption 1. *One of the following two conditions holds:*

(i) $\text{rank}(\Sigma^{+}_{\Lambda\Psi}) > r_a$;

(ii) $\rho_\ell(\Sigma_F \Sigma_\Lambda) \neq \rho_\ell(\Sigma_{\overline{F}} \Sigma_\Psi)$ *for some* $\ell \leq r_a$.

Here, $\rho_\ell(A)$ denotes the ℓ-th largest eigenvalue of a square matrix A, and $\text{rank}(A)$ denotes the rank of an $m \times n$ matrix A, which refers to the number of pivot columns in an echelon of A. Here, pivot columns refer to the columns which have entries distinct from 0.

3.2 Shrinkage Estimation

Shrinkage estimation is a statistical algorithm which aims to stabilize the estimates, reduce errors, and smooth fluctuations by narrowing the range of extreme values towards the sample mean (Xie et al. 2016)[33]. We also refer to Nkurunziza et al. (2016)[25] for more details about shrinkage and LASSO estimators. In this section, we introduce the purpose of using shrinkage estimator and how it works for our research. We also explain how to obtain the shrinkage estimator by achieving the minimization of the penalized least square (PLS) functions in details.

First we begin with introducing some special cases of the factor model. After the break point, if the loadings of the pre-break factors remain the same, then $\Gamma_1^0 = 0$; while if no new factors appear, then $\Gamma_2^0 = 0$. Suppose that the post-break factor loadings is $\Psi^0 = (\Lambda^0 + \Gamma_1^0, \Gamma_2^0)$, which represents the loadings before and after break point, then the model in (2.2) can be rewritten as

$$X_b = F_b \Psi^{0\prime} + e_b. \tag{3.2}$$

In (2.1) and (3.2), the factors and their loadings have to be estimated together. So the normalization restrictions are applied to estimate the factor model (Cheng et al. 2016)[14]. The factor model in (2.1) and (3.2) can be rewritten as

$$\begin{aligned} X_a &= (F_a R_a)(R_a^{-1} \Lambda^{0\prime}) + e_a = F_a^R \Lambda^{R\prime} + e_a, \\ X_b &= (F_b R_b)(R_b^{-1} \Psi^{0\prime}) + e_b = F_b^R \Psi^{R\prime} + e_b, \end{aligned} \tag{3.3}$$

where $F_a^R = F_a R_a$ and $F_b^R = F_b R_b$ and R_a and R_b are transformation matrices.

Suppose that $r_a + r_b \leq k$ where k is an upper bound of the number of factors. The factor model in (3.3) is rewritten as the following augmented system (Cheng et al. 2016)[14].

$$\begin{aligned} X_a &= \begin{bmatrix} F_a^R & F_{a,1}^{R\perp} & F_{a,2}^{R\perp} \end{bmatrix} \begin{bmatrix} \Lambda^{R\prime} \\ \mathbf{0}_{(r_b - r_a)\times N} \\ \mathbf{0}_{(k-r_b)\times N} \end{bmatrix} + e_a = F_a^{R+} (\Lambda^{R+})' + e_a, \\ X_b &= \begin{bmatrix} F_{b,1}^R & F_{b,2}^R & F_b^{R\perp} \end{bmatrix} \begin{bmatrix} \Lambda^{R\prime} + \Gamma_1^{R\prime} \\ \Gamma_2^{R\prime} \\ \mathbf{0}_{(k-r_b)\times N} \end{bmatrix} + e_b = F_b^{R+} (\Lambda^{R+} + \Gamma^{R+})' + e_b. \end{aligned} \tag{3.4}$$

The matrix $F_a^{R\perp}$ is an orthogonal complement of F_a^R, and it can be partitioned into two sub-matrices ${F_{a,1}}^{R\perp} \in R^{T_0\times(r_b-r_a)}$ and $F_{a,2}^{R\perp} \in R^{T_0\times(k-r_b)}$. Similarly, $F_b^{R\perp}$ denotes an orthogonal complement of $F_b^R \in R^{T_b\times r_b}$ and F_b^R can be partitioned into two sub-matrices $F_{b,1}^R$ and $F_{b,2}^R$. Here, we treat F_a^R and F_b^R as the true factors and $F_a^{R\perp}$ and $F_b^{R\perp}$ as the irrelevant factors. In the augmented system (3.4), we obtain Λ^{R+} and $(\Lambda^{R+}+\Gamma^{R+})$, which are the factor loadings for pre- and post-break periods respectively. They will be further analyzed by the shrinkage estimator.

The numbers of factors r_a and r_b and instability in factor loadings can be estimated concurrently in model (3.4) by selecting the zero and nonzero components in Λ^{R+} and Γ^{R+} respectively. Therefore, it is essential to find such estimators, which turn out to be the shrinkage estimators, that can differentiate zero and nonzero components of matrices Λ^{R+} and Γ^{R+}. In the rest of this section, the whole procedure will be explained in details.

Since the break date is known, we obtain the shrinkage estimator by minimizing the penalized least square (PLS) function which consists of two group-LASSO penalty functions (Yuan and Lin, 2006)[34].

First, we estimate k potential factors. It is determined by the principle component estimators in each subsample. For instance, suppose $j \in \{a, b\}$, let $\widetilde{F}_j$ be a $T_j \times k$ matrix which consists of the orthonormalized eigenvectors of $(NT_j)^{-1}X_jX_j'$ with the first k largest eigenvalues. Since $\widetilde{F}_a$ and $\widetilde{F}_b$ are known, the shrinkage estimators $(\widehat{\Lambda}, \widehat{\Gamma})$ of the factor loadings $(\Lambda^{R+}, \Gamma^{R+})$ can be achieved by minimizing the following PLS criterion function:

$$(\widehat{\Lambda}, \widehat{\Gamma}) = \underset{\Lambda \in R^{N\times k}, \Gamma \in R^{N\times k}}{\operatorname{argmin}} [M(\Lambda, \Gamma) + P_1(\Lambda) + P_2(\Gamma)], \tag{3.5}$$

where

$$\begin{aligned} M(\Lambda, \Gamma) &= (NT)^{-1}\left[\left\|X_a - \widetilde{F}_a\Lambda'\right\|^2 + \left\|X_b - \widetilde{F}_b(\Lambda+\Gamma)'\right\|^2\right], \\ P_1(\Lambda) &= \alpha_{NT}\sum_{\ell=1}^{k}\omega_\ell^\lambda\|\Lambda_\ell\| \text{ and } P_2(\Gamma) = \beta_{NT}\sum_{\ell=1}^{k}\omega_\ell^\gamma\|\Gamma_\ell\|, \end{aligned} \tag{3.6}$$

where the ℓ subscript of Λ_ℓ and Γ_ℓ denotes the ℓ-th column of matrices Λ and Γ, α_{NT} and β_{NT} stand for positive constant parameters varying with N and T, ω_ℓ^λ and ω_ℓ^γ are data-adapting weights defined as follow

$$\begin{aligned} \omega_\ell^\lambda &= \left(N^{-1}\|\widetilde{\Lambda}_\ell\|^2\mathcal{I}_{\{\widetilde{\Lambda}_\ell\neq 0_{N\times 1}\}} + N^{-1}\|\widetilde{\Lambda}_{\ell,LS}\|^2\mathcal{I}_{\{\widetilde{\Lambda}_\ell = 0_{N\times 1}\}}\right)^2, \\ \omega_\ell^\gamma &= \left(N^{-1}\|\widetilde{\Gamma}_\ell\|^2\mathcal{I}_{\{\widetilde{\Gamma}_\ell\neq 0_{N\times 1}\}} + N^{-1}\|\widetilde{\Gamma}_{\ell,LS}\|^2\mathcal{I}_{\{\widetilde{\Gamma}_\ell = 0_{N\times 1}\}}\right)^2. \end{aligned} \tag{3.7}$$

Those two weights in (3.7) are constructed to distinguish between the zero and nonzero columns of Λ^{R+} and Γ^{R+}. Here, $\widetilde{\Lambda} \in R^{N\times K}$ and $\widetilde{\Gamma} \in R^{N\times K}$ denote some preliminary estimators of factor loadings Λ^{R+} and Γ^{R+}; $\widetilde{\Lambda}_{LS} = T_a^{-1}X_a'\widetilde{F}_a$ and $\widetilde{\Psi}_{LS} = T_b^{-1}X_b'\widetilde{F}_b$ are two unrestricted least square estimators of the pre- and post-break factor loading matrices. The change of factor loadings $\widetilde{\Gamma}_{LS} = \widetilde{\Psi}_{LS} - \widetilde{\Lambda}_{LS}$.

Note that, in this case, the simplest preliminary estimators are the unrestricted least squares estimators $\widetilde{\Lambda}_{LS}$ and $\widetilde{\Lambda}_{LS}$. We apply $\widetilde{\Lambda}_{LS}$ and $\widetilde{\Lambda}_{LS}$ if other preliminary estimators $\widetilde{\Lambda}$ and $\widetilde{\Gamma}$ have zero columns, which happens sometimes. Therefore we include

$\mathcal{I}_A(x)$ in (3.7) which denotes the indicator function of the subset A, i.e. it takes the value 1 if $x \in A$, otherwise it takes the value 0.

The penalty functions $P_1(\Lambda)$ and $P_2(\Gamma)$, defined in terms of the column norms $\|\Lambda_\ell\|$ and $\|\Gamma_\ell\|$, are group-LASSO penalties (Yuan and Lin, 2006)[34]. We use this group-LASSO penalties to handle large-scale factor models and eliminate irrelevant factors which have zero factor loadings for all series (Cheng et al. 2016)[14]. It either sets all the elements in a group equal to zero or estimates them as non-zeros altogether.

3.3 Estimators of the Number of Factors and the Break Point

In this section, we show that by using the shrinkage estimators introduced in Section 3.2, the numbers of pre- and post-break factors r_a and r_b and the existence of the break point $\mathcal{B}_0$ can be specified simultaneously. The structural instability can be identified as well. We use a binary variable $\mathcal{B}_0 \in \{0, 1\}$ to indicate the occurrence of the break point. $\mathcal{B}_0$=0 if there is no break, otherwise $\mathcal{B}_0$=1. Thus, the structural instability in (2.3) can be rewritten as $\mathcal{B}_0$=1 and $r_a = r_b$.

The estimator of $\mathcal{B}_0$ is defined as

$$\widehat{\mathcal{B}} = \mathcal{I}_{\{\|\widehat{\Gamma}\|>0\}}. \tag{3.8}$$

To show that the estimator $\widehat{\mathcal{B}} = 0$, it is equivalent to show that $\Gamma^0 = (\Gamma_1^0, \Gamma_2^0) = 0$ in

(2.2). This holds if and only if $\Gamma^R = 0$, where $\Gamma^R = (\Gamma_1^R, \Gamma_2^R)$ is obtained by rewriting the normalized version of the post-break factor model in (3.3) as

$$X_b = F_b^R \Psi^{R\prime} + e_b = F_{b,1}^R (\Lambda^R + \Gamma_1^R)' + F_{b,2}^R \Gamma_2^{R\prime} + e_b, \tag{3.9}$$

where $F_b^R = (F_{b,1}^R, F_{b,2}^R)$, $\Psi^R = (\Lambda^R + \Gamma_1^R, \Gamma_2^R)$ and Γ^R=(Γ_1^R,Γ_2^R).

The estimators of r_a and r_b are given by the numbers of the last non-zero columns of $\widehat{\Lambda}$ and $\widehat{\Gamma}$ respectively, which are

$$\begin{aligned} \widehat{r}_a &= \min \{j \geq 1 : \|\widehat{\Lambda}_\ell\| = 0 \text{ for all } \ell > j\}, \\ \widehat{r}_b &= \max \left(\hat{r}_a, \ \min \{j \geq 1 : \|\widehat{\Gamma}_\ell\| = 0 \text{ for all } \ell > j\} \right). \end{aligned} \tag{3.10}$$

3.4 Post Model Selection Estimation

In the previous sections, we obtain the estimators $\widehat{r}_a$, $\widehat{r}_b$ and $\widehat{\mathcal{B}}$ from the shrinkage estimator. In this section, we demonstrate that the shrinkage estimator also contributes to the estimation of the loading matrices Λ and Γ. When we estimate the non-zero coefficients of the loading matrices, we have to re-estimate the loadings using least squares conditional on the selected $\widehat{r}_a$, $\widehat{r}_b$ and $\widehat{\mathcal{B}}$. In this case, we can optimize the penalty terms of the shrinkage estimator. The obtained estimator is called the post model selection (PMS) estimator.

First, consider there is no structural break ($\widehat{\mathcal{B}} = 0$), then we can re-estimate the factor model on the full sample. Suppose that $\widetilde{F} \in R^{T\times k}$ is an orthonormalized matrix which consists of the first k principal components of the full sample. Let $\overline{\Lambda}$ denote the first $\widehat{r}_a$ columns of the full sample least squares estimators $\widetilde{\Lambda}_{LS} = T^{-1}X'\widetilde{F} = \widetilde{\Psi}_{LS}$ since there is no break. Thus, we set $\overline{\Lambda} = \overline{\Psi}$.

Second, if there exists a break ($\widehat{\mathcal{B}} = 1$), then we need to re-estimate the two subsamples of factors and loadings separately. Let $\widetilde{F}_a$ and $\widetilde{F}_b$ denote those two subsamples. Let $\overline{\Lambda}$ denote the first $\widehat{r}_a$ columns of $\widetilde{\Lambda}_{LS} = T^{-1}X_a'\widetilde{F}_a$. Similarly, let $\overline{\Psi}$ denote the first $\widehat{r}_b$ columns of $\widetilde{\Psi}_{LS} = T^{-1}X_b'\widetilde{F}_b$. The PMS estimators are defined as

$$\widehat{\Lambda}_{PMS} = (\overline{\Lambda}, \mathbf{0}) \text{ and } \widehat{\Psi}_{PMS} = (\overline{\Psi}, \mathbf{0}), \tag{3.11}$$

where $\mathbf{0}$ is the zero matrix.

3.5 Estimation of the Tuning Parameters

In this section, we focus on the calculation of the shrinkage estimator. We propose a method to obtain the penalty functions by choosing the most appropriate tuning parameters α_{NT} and β_{NT}.

The group LASSO penalty functions $P_1(\Lambda)$ and $P_2(\Gamma)$ are determined by the tuning parameters α_{NT} and β_{NT}. They are the weights with respect to X_a and X_b. The tuning parameters are defined as

$$\alpha_{NT} = \kappa_1 N^{-1/2} C_{NT_a}^{-3} \text{ and } \beta_{NT} = \kappa_2 N^{-1/2} C_{NT_b}^{-3}, \tag{3.12}$$

where $C_{NT_a} = \min\,(N^{1/2}, T_a^{1/2})$, and $C_{NT_b} = \min\,(N^{1/2}, T_b^{1/2})$. Let κ_1 and κ_2 be two constant numbers equal to

$$\begin{aligned} \kappa_1 &= c_1 \Big\{ (NT_a)^{-1/2} \Big\| e_a(\widetilde{\Lambda}) \Big\| + (NT_b)^{-1/2} \Big\| e_b(\widetilde{\Lambda} + \widetilde{\Gamma}) \Big\| \Big\}, \\ \kappa_2 &= c_2 (NT_b)^{-1/2} \Big\| e_b(\widetilde{\Lambda} + \widetilde{\Gamma}) \Big\|, \end{aligned} \tag{3.13}$$

where $\widetilde{\Lambda}$ and $\widetilde{\Gamma}$ are preliminary estimators, $e_a(\Lambda)$ and $e_b(\Lambda + \Gamma)$ are residual matrices defined as

$$e_a(\Lambda) = X_a - \widetilde{F}_a \Lambda' \text{ and } e_b(\Lambda + \Gamma) = X_b - \widetilde{F}_b (\Lambda + \Gamma)'. \tag{3.14}$$

In (3.12), C_{NT_a} and C_{NT_b} are rates related to N and T. We can balance these two rates by varying α_{NT} and β_{NT} and replace the whole sample size T by the subsamples T_a and T_b. We set up a default value for the constants c_1 and c_2 as $c_1 = c_2 = 1$, but we also apply a cross validation procedure to calibrate c_1 and c_2 over a fixed interval in finite samples.

3.6 Two-Step Estimation Algorithm

In this section, we introduce an effective two-step estimation algorithm from Cheng et al. (2016)[14]. First, we obtain a preliminary estimator, then we use it to adjust the penalty terms of the second-step shrinkage estimator.

This procedure develops the finite sample performance with respect to two aspects.

First, the residual matrices in (3.13) are optimized when $\widetilde{\Lambda}$ and $\widetilde{\Gamma}$ are based on a first-step estimator. Thus, the tuning parameters α_{NT} and β_{NT} are improved in the second step. Second, the preliminary estimator $\widetilde{\Gamma}$ which is obtained through a rotation of the factor loadings Λ^R and Ψ^R also has better performance.

Let $\widetilde{\Lambda}^{(i)}$, $\widetilde{\Psi}^{(i)}$ and $\widetilde{\Gamma}^{(i)}$ denote the preliminary estimators with Step i, $i = 1$ and 2. Let $\widehat{\Lambda}^{(i)}$, $\widehat{\Psi}^{(i)}$ and $\widehat{\Gamma}^{(i)}$ denote the penalty least square (PLS) estimators in Step i. Let $\widehat{\Lambda}^{(i)}_{PMS}$, $\widehat{\Gamma}^{(i)}_{PMS}$ and $\widehat{\Psi}^{(i)}_{PMS}$ denote the post model selection (PMS) estimators in step i. The two-step estimation procedure is as follows:

Algorithm 1 (Two-Step Estimation Procedure)

1. First-Stage Shrinkage Estimation:

   1.1. Compute the unrestricted least square estimators $\widetilde{\Lambda}_{LS}$ and $\widetilde{\Gamma}_{LS}$.

   1.2. Let $\widetilde{\Lambda}^{(1)}$=$\widetilde{\Lambda}_{LS}$ and $\widetilde{\Gamma}^{(1)}$=$\widetilde{\Gamma}_{LS}$ be the first preliminary estimators. Calculate ω_ℓ^λ, ω_ℓ^γ, α_{NT} and β_{NT} from (3.7), (3.12) and (3.13) with $\widetilde{\Lambda} = \widetilde{\Lambda}^{(1)}$ and $\widetilde{\Gamma} = \widetilde{\Gamma}^{(1)}$.

   1.3. Compute the shrinkage estimators $\widetilde{\Lambda}^{(1)}$ and $\widetilde{\Gamma}^{(1)}$ by minimizing the criterion function in (3.5).

   1.4. Estimate r_a and r_b from (3.10) with $\widehat{\Lambda} = \widehat{\Lambda}^{(1)}$ and $\widehat{\Gamma} = \widehat{\Gamma}^{(1)}$. Name the estimators as $\widehat{r}_a^{(1)}$ and $\widehat{r}_b^{(1)}$.

   1.5. Construct the PMS estimators $\widehat{\Lambda}^{(1)}_{PMS}$ and $\widehat{\Psi}^{(1)}_{PMS}$ by (3.11). If $\widehat{r}_a^{(1)} = \widehat{r}_b^{(1)}$, then we define the rotation of the columns of $\overline{\Psi}^{(1)}$ as

$$\overline{\Psi}^{(1)}_R = \overline{\Psi}^{(1)} Q, \tag{3.15}$$

where Q=VU' and let UDV' denote the singular value decomposition of $\overline{\Lambda}^{(1)'}\overline{\Psi}^{(1)}$ such that $\overline{\Lambda}^{(1)'}\overline{\Psi}^{(1)} = UDV'$. The modified PMS estimator of Ψ is defined as

$$\widehat{\Psi}^{(1)}_{PMS-R} = \left(\overline{\Psi}^{(1)}_{R}, \mathbf{0}_{\Psi^{(1)}}\right) \in R^{N\times k}. \tag{3.16}$$

2. Second-Stage Shrinkage Estimation

2.1. Let

$$\begin{aligned} \widetilde{\Lambda}^{(2)} &= \widehat{\Lambda}^{(1)}_{PMS}, \\ \widetilde{\Psi}^{(2)} &= \begin{cases} \widehat{\Psi}^{(1)}_{PMS-R} \;\; if\; \widehat{r}^{(1)}_b = \widehat{r}^{(1)}_a \\ \widehat{\Psi}^{(1)}_{PMS} \;\; if\; \widehat{r}^{(1)}_b > \widehat{r}^{(1)}_a, \end{cases} \\ \widetilde{\Gamma}^{(2)} &= \widetilde{\Psi}^{(2)} - \widetilde{\Lambda}^{(2)}, \end{aligned} \tag{3.17}$$

and ω^{λ}_{ℓ},ω^{γ}_{ℓ}, α_{NT}, and β_{NT} are calculated from (3.7), (3.12) and (3.13) with $\widetilde{\Lambda}$=$\widetilde{\Lambda}^{(2)}$ and $\widetilde{\Gamma}$=$\widetilde{\Gamma}^{(2)}$.

2.2. Compute the shrinkage estimators $\widehat{\Lambda}^{(2)}$ and $\widehat{\Gamma}^{(2)}$ by minimizing the criterion function (3.5).

2.3. Compute $\mathcal{B}^{(2)}_0$, $\widehat{r}^{(2)}_a$, and $\widehat{r}^{(2)}_b$ from (3.10) with $\widetilde{\Lambda}$=$\widetilde{\Lambda}^{(2)}$ and $\widetilde{\Gamma}$=$\widetilde{\Gamma}^{(2)}$.

2.4. Construct the PMS estimator $\widehat{\Lambda}^{(2)}_{PMS}$ and $\widehat{\Psi}^{(2)}_{PMS}$ by the definition in (4.9) given the selection of the estimators $\widehat{\mathcal{B}}^{(2)}$, $\widehat{r}^{(2)}_a$ and $\widehat{r}^{(2)}_b$.

The PMS estimator of the first stage is used to improve the preliminary estimators of the second stage. In Step 1.5, we obtain a transformation of $\overline{\Psi}^{(1)}$ as $\overline{\Psi}^{(1)}_R$ since the rotated matrix performs better in locating the structural break when there is no instability in the model. The transformation does not affect the procedure of the asymptotic theory.

To give more explanation of Step 1.5, when we do the transformation, first we need to generate an orthogonal matrix Q which can minimize $\|\overline{\Lambda}^{(1)} - \overline{\Psi}^{(1)}Q\|$. It suffices to maximize the correlation between the columns of $\overline{\Lambda}^{(1)}$ and $\overline{\Psi}^{(1)}Q$. The solution is $Q = VU'$ from the previous literature of Schönemann (1966)[28]. In case of Step 1.5, $\overline{\Lambda}^{(1)'}\overline{\Psi}^{(1)} = UDV'$ by the singular value decomposition.

3.7 Cross Validation

We adopt the cross validation method (Cheng et al. 2016)[14] to calibrate the constant $c = (c_1, c_2) \in \mathcal{C}$ which appears in the penalty weights in (3.13). It has default value $c_1 = c_2 = 1$. We have already partitioned the sample according to T time points. Since we are not able to observe the factors in current environment, we also need to subdivide the sample cross-sectionally.

First, the data in the cross-sectional dimension creates disjoint subsamples $X_{(-jN)}$ (N-regression) and X_{jN} (N-prediction). Then, the procedure in Section 3.3 is applied to this subsample $X_{(-jN)}$. This produces the estimator of the unobserved factors and the estimators $\widehat{r_a}(-jN, c)$, $\widehat{r_b}(-jN, c)$ and $\widehat{\mathcal{B}}(-jN, c)$ for a given value c. We then partition the left subsample X_{jN} along the T dimension into regression and prediction samples. If the structural break is taken place in the partitioned subsample, then the regression and prediction samples need to be treated separately for the pre- and post-break periods. Given the estimators of the factors and the loadings, we can then generate sample forecasts for the prediction sample.

The cross-validation criterion used in this major paper is built on mean-squared forecast errors (MSFE). The tuning constants are chosen to minimize the MSFE for given c. The minimization is accomplished over a bounded set $\mathcal{C}$. The adjusted constant improves the performance of the finite sample.

3.8 Asymptotic Theory

In this section, we study the asymptotic performance of the PLS estimator. To estimate the number of factors from the observed data, the penalty for overfitting must be a function of both the cross-section dimension N and the time dimension T (Bai and Ng, 2002)[5]. In the research of Lewbel (1991)[21] and Donald (1997)[15], they assume either N or T is fixed. The reason is that the functions of N and T, the AIC and BIC, usually do not work when both dimensions are large. However, in this major paper, we assume that both N and T converge to infinity. This adjustment is necessary because of our empirical knowledge. The time dimension of the dataset is not able to be assumed as a fixed T since it is too large (Bai and Ng, 2002)[5]. Moreover, the cross-section dimension is larger relative to the time dimension. We present some assumptions and state some theorems, which are mainly from Cheng et al. (2016)[14], on the large sample properties of the preliminary estimators $\widetilde{\Lambda}$ and $\widetilde{\Gamma}$ and on the convergence rates of the sequences α_{NT} and β_{NT} below.

First, Theorem 3.1 and 3.2 give the stochastic order of the preliminary estimators $\widetilde{\Lambda}$ and $\widetilde{\Gamma}$ which may result in the change of the data-dependent weights ω_ℓ^λ and ω_ℓ^γ defined in (3.7). Their columns are separated into two parts, the first part is from 1 to the number of pre-break factors r_a and the second part is from the number of post-break factors r_b to k. Let O_p (Big O) denote the stochastic boundedness. For example, $X_n = O_p(a_n)$ means that X_n/a_n is stochastically bounded. Specifically, for any $\epsilon > 0$, there exists a finite $M > 0$ and a finite $N > 0$ such that $\forall n > N$, $\mathrm{P}(|X_n/a_n| > M) < \epsilon$.

Theorem 3.1. *If $N, T \to \infty$ with $\sqrt{N}/T \to \infty$, the preliminary estimators $\widetilde{\Lambda}$ and $\widetilde{\Gamma}$ satisfy*

(i) $\lim_{N,T\to\infty} \Pr\big(N^{-1}\|\widetilde{\Lambda}_\ell\|^2 \geq C\big) = 1$ *for* $\ell = 1, \cdots, r_a$,
$N^{-1}\|\widetilde{\Lambda}_\ell\|^2 = O_p(C_{NT}^{-2})$ *for* $\ell = r_a + 1, \cdots, k$;
(ii) If $\Gamma^0 \neq 0$, $\lim_{N,T\to\infty} \Pr\big(N^{-1}\|\widetilde{\Gamma}_\ell\|^2 \geq C\big) = 1$ *for* $\ell = 1, \cdots, r_b$,
$N^{-1}\|\widetilde{\Gamma}_\ell\|^2 = O_p(C_{NT}^{-2})$ *for* $\ell = r_b + 1, \cdots, k$;
(iii) If $\Gamma^0 = 0$, $N^{-1}\|\widetilde{\Gamma}_\ell\|^2 = O_p(C_{NT}^{-2})$ *for* $\ell = 1, \cdots, k$. □

For the proof of this Theorem, we refer to Cheng et al. (2016)[14]. Here, the large penalties generate shrinkage estimators that equal to 0 with probability approaching to 1. Also, $\widetilde{\Lambda}_\ell = 0$ is a special case of $N^{-1}\|\widetilde{\Lambda}_\ell\|^2 = O_p(C_{NT}^{-2})$, and so as $\widetilde{\Gamma}_\ell$.

Theorem 3.2. *If $N, T \to \infty$ with $\sqrt{N}/T \to \infty$, the preliminary estimators $\widetilde{\Lambda}_{LS}$ and $\widetilde{\Gamma}_{LS}$ satisfy*

(i) $\lim_{N,T\to\infty} \Pr\big(N^{-1}\|\widetilde{\Lambda}_{LS,\ell}\|^2 \geq C\big) = 1$ *for* $\ell = 1, \cdots, r_a$,
$N^{-1}\|\widetilde{\Lambda}_{LS,\ell}\|^2 = O_p(C_{NT}^{-2})$ *for* $\ell = r_a + 1, \cdots, k$;
(ii) If $\Gamma^0 \neq 0$, $\lim_{N,T\to\infty} \Pr\big(N^{-1}\|\widetilde{\Gamma}_{LS,\ell}\|^2 \geq C\big) = 1$ *for* $\ell = 1, \cdots, r_b$,

$N^{-1}\|\widetilde{\Gamma}_{LS,\ell}\|^2 = O_p(C_{NT}^{-2})$ *for* $\ell = r_b + 1, \cdots, k$;

(iii) If $\Gamma^0 = 0$, $N^{-1}\|\widetilde{\Gamma}_{LS,\ell}\|^2 = O_p(C_{NT}^{-2})$ *for* $\ell = 1, \cdots, k$. $\square$

The proof of this Theorem follows from the results in Cheng et al. (2016)[14]. Theorem 3.2 takes care of the situation that $\widetilde{\Lambda}$ or $\widetilde{\Gamma}$ has zero columns. In this case, ω_ℓ^λ and ω_ℓ^γ depend on $\widetilde{\Lambda}_{LS}$ and $\widetilde{\Lambda}_{LS}$.

Then, the following Assumptions 2 and 3 are designed for the large sample behaviour of the factors and they correspond to Assumptions A and B from Cheng et al. (2016)[14]. For the post-break period, let $\overline{F}_t^0 = (F_t^{0\prime}, F_t^{*\prime})' \in \mathbb{R}^{r_b}$ denote the r_b factors after the break and $C \in \mathbb{R}$ denotes a generic positive constant.

Assumption 2. —***Factors:***
$\mathbb{E}\big[\|F_t^0\|^4\big] \leq C$, $\mathbb{E}\big[\|\overline{F}_t^0\|^4\big] \leq C$ *and there exist positive definite nonrandom matrices* Σ_F *and* $\Sigma_{\overline{F}}$ *not depending on* N *such that* $T_0^{-1}\sum_{t=1}^{T_0} F_t^0 F_t^{0\prime} = \Sigma_F + O_p(T_0^{-1/2})$ *and* $T_1^{-1}\sum_{t=T_0+1}^{T} \overline{F}_t^0 \overline{F}_t^{0\prime} = \Sigma_{\overline{F}} + O_p(T_1^{-1/2})$. $\square$

Let $\Lambda^0 = (\lambda_1^0, \cdots, \lambda_N^0)'$, where $\lambda_i^0 \in R^{r_a \times 1}$ is the factor loadings for $i = 1, \cdots, N$ before the break. Similarly, let $\Psi^0 = (\psi_1^0, \cdots, \psi_N^0)'$, where $\psi_i^0 \in R^{r_b \times 1}$ is the factor loadings after the break.

Assumption 3. —***Factor Loadings:***
(1) $\|\lambda_i^0\| \leq C$, $\|\psi_i^0\| \leq C$ *and there exist nonrandom matrices* Σ_Λ, Σ_Ψ *and* $\Sigma_{\Lambda\Psi}$ *not depending on* N *such that* $\|\Lambda^{0\prime}\Lambda^0/N - \Sigma_\Lambda\| \to 0$, $\|\Psi^{0\prime}\Psi^0/N - \Sigma_\Psi\| \to 0$ *and* $\|\Lambda^{0\prime}\Psi^0/N - \Sigma_{\Lambda\Psi}\| \to 0$ *as* $N \to \infty$, *where* Σ_Λ *and* Σ_Ψ *are positive definite.*
(2) The matrices $\Sigma_\Lambda\Sigma_F$ *and* $\Sigma_\Psi\Sigma_{\overline{F}}$ *both have distinct eigenvalues.* $\square$

Assumption 3 guarantees that each factor has a nontrivial contribution to the variance of X_{it}. We only consider nonrandom factor loadings.

Next, suppose that $T_0/T \to \tau_0$ for some constant $\tau_0 \in (0,1)$ as $T \to \infty$. Let $e = [e_1, \cdots, e_T] \in R^{N\times T}$ be the matrix of errors and e_{it} denote the (i,t) element of series i in t. Below, we present Assumption 4 which restricts the high dependence between the time dimensional partitions and the cross-sectional partitions in the error terms. Meanwhile, we also need Assumption 5 to restrict high dependence between the factors and the error terms.

Assumption 4. —***Time and Cross-Section Dependence:***

(i) $\mathbb{E}[e_it] = 0,\ \mathbb{E}\big[|e_{it}|\big] \le C;$

(ii) $\mathbb{E}\big[N^{-1}\sum_{i=1}^N e_{is}e_{it}\big] = \sigma_N(s,t),\ \big|\sigma_N(s,s)\big| \le C \textit{ for all } s,$

$T^{-1}\sum_{s=1}^T\sum_{t=1}^T \big|\sigma_N(s,t)\big| \le C;$

(iii) $\mathbb{E}[e_{it}e_{jt}] = \tau_{ij,t} \textit{ with } |\tau_{ij,t}| \le |\tau_{ij}| \textit{ for some } \tau_{ij} \textit{ and for all } t, \textit{ and}$

$N^{-1}\sum_{i=1}^N\sum_{j=1}^N \big|\sigma_N(s,t)\big| \le C;$

(iv) $\mathbb{E}[e_{it}e_{js}] = \tau_{ij,ts} \textit{ and } (NT)^{-1}\sum_{i=1}^N\sum_{j=1}^N\sum_{t=1}^N\sum_{s=1}^N |\tau_{ij,ts}| \le C;$

(v) For every $(t,s),\ \mathbb{E}\left[\left|N^{-1/2}\sum_{i=1}^N \big[e_{is}e_{it} - \mathbb{E}[e_{is}e_{it}]\big]\right|^4\right] \le C;$

(vi) $\rho_1\big((NT)^{-1}e_a e_a'\big) = O_p\big(max[N^{-1}, T^{-1}]\big)$ *and*

$\rho_1\big((NT)^{-1}e_b e_b'\big) = O_p\big(max[N^{-1}, T^{-1}]\big).$ □

Assumption 5. —***Weak Dependence between Factors and Errors:***

$\mathbb{E}\left[N^{-1}\sum_{i=1}^N \left\|T^{-1/2}\big(\sum_{t=1}^{T_0} F_t^0 e_{it} + \sum_{t=T_0+1}^T \overline{F}^0_t e_{it}\big)\right\|^2\right] \le C.$ □

These two assumptions are analogous to the Assumptions C and D in Bai and Ng (2002)[5]. Note that Assumption 4(vi) is proposed as a condition to select the number of factors (Amengual and Watson, 2007)[3].

Moreover, we need another assumption for the tuning parameter α_{NT} and β_{NT}, which determine the overall penalization. The following Assumption 6 states the convergence rate C_{NT} of the unrestricted least square estimator at which the tuning parameters α_{NT} and β_{NT} vanish asymptotically. As in Bai and Ng (2002)[5], we set $C_{NT} = \min(T^{1/2},\ N^{1/2})$, where C_{NT} is the convergence rate of the unrestricted least square estimator.

Assumption 6. *The tuning parameters α_{NT} and β_{NT} satisfy*

(i) $\alpha_{NT} = O(N^{-1/2}C_{NT}^{-1})$ and $\beta_{NT} = O(N^{-1/2}C_{NT}^{-1})$;

(ii) $N^{-1/2}C_{NT}^{-5} = o(\alpha_{NT})$ and $N^{-1/2}C_{NT}^{-5} = o(\beta_{NT})$. $\square$

In Assumption 6, the boundaries of α_{NT} and β_{NT} limit the extent of the overall penalization of shrinkage estimators. Assumption 6(i) ensures that the penalties on the nonzero columns are small when the weights ω_ℓ^λ and ω_ℓ^γ are stochastically bounded. Also, we propose to shrink the estimators of zero columns to zero. In Assumption 6(ii), we introduce the lower bound which requires the tuning parameter α_{NT} and β_{NT} converge to zero slowly.

Finally, we present Theorem 3.3 which gives the asymptotic limits of the PLS estimators $\widehat{\Lambda}$ and $\widehat{\Gamma}$ as the following Theorem 3.3.

Theorem 3.3. *Suppose that Assumptions 2-6 holds, then*

(i) Pre-break loadings of relevant factors: $N^{-1}\|\widehat{\Lambda}_\ell - \Lambda_\ell^R\|^2 = O_p(C_{NT}^{-2})$ for $\ell = 1, \cdots, r_a$;

(ii) Pre-break loadings of irrelevant factors: $\lim\limits_{N,T\to\infty} \Pr\big(\|\widehat{\Lambda}_\ell\|^2 = 0\big) = 1$

for $\ell = r_a + 1, \cdots, k$;

(iii) Post-break changes in loadings of relevant factors: If $\Gamma^0 \neq 0$,

$N^{-1}\|\widehat{\Gamma}_\ell - \Gamma_\ell^R\|^2 = O_p(C_{NT}^{-2})$ *for* $\ell = 1, \cdots, r_b$;

(iv) No-break: If $\Gamma^0 = 0$, $\lim_{N,T\to\infty} \Pr\big(\|\widehat{\Gamma}_\ell\|^2 = 0\big) = 1$ *for* $\ell = 1, \cdots, r_b$;

(v) Post-break changes in loadings of irrelevant factors:

$\lim_{N,T\to\infty} \Pr\big(\|\widehat{\Gamma}_\ell\|^2 = 0\big) = 1$ *for* $\ell = r_b + 1, \cdots, k$. $\square$

The proof of Theorem 3.3 follows from the results in Cheng et al. (2016)[14]. Here, Theorem 3.3(i) and (ii) show that the factor loadings of the irrelevant factors are turned out to be 0 with probability approaching 1 because of the penalization. For $\ell = 1, \cdots, r_a$, the PLS estimators $\widehat{\Lambda}_\ell$ and $\widehat{\Gamma}_\ell$ converge in probability to the factor loadings Λ_ℓ^R and Γ_ℓ^R in the transformed factor model (3.3) respectively. Theorem 3.3(iii) to (v) are used to detect the structural instability by the asymptotic properties of the PLS estimators of the changes in the factor loadings. If there is no structural instability, like in part (iv), the PLS estimators of the changes $\widehat{\Gamma}_\ell$ are turned out to have columns from 1 to r_b equal to 0 with probability approaching to 1. If there does exists a structural instability, only the estimators $\widehat{\Gamma}_\ell$ such that with column $\ell = r_b + 1, \cdots, k$ satisfies part (v).

To sum up, Theorem 3.3 states that the factor loadings of irrelevant factors can be estimated as 0 w.p.a 1. Meanwhile, the changes in the loadings of the relevant factors can be estimated as 0 with probability approaching to 1 if their factor loadings are not in the control of the structural instability. An equivalent way to get the PLS estimator is to show that the asymptotic limits of $N^{-1}\|\Lambda_\ell^R\|^2$ and $N^{-1}\|\Gamma_\ell^R\|^2$ in Part (i) and (iii) are bounded away from 0 which can be obtained from Assumption 1. The following Theorem 3.4 displays the asymptotic result of the estimators.

Theorem 3.4. *Suppose that Assumptions 1-6 hold, then the following hold*

$$\lim_{N,T\to\infty} \Pr(\widehat{\mathcal{B}} = \mathcal{B}_0) = 1, \lim_{N,T\to\infty} \Pr(\widehat{r_a} = r_a) = 1, \lim_{N,T\to\infty} \Pr(\widehat{r_b} = r_b) = 1.$$

Theorem 3.4 shows that the model can be estimated simultaneously for any set of preliminary estimators $\widetilde{\Lambda}$ and $\widetilde{\Gamma}$ that satisfy Theorems 3.1 and 3.2. Theorem 3.4 is given in Cheng et al. (2016)[14]. Furthermore, for the convenience of the reader, we also present a detailed proof in the appendix.

Given the existence of the structural instability, we obtain a set of columns as following,

$$\mathcal{Z} = \Big\{ \ell : N^{-1}\|\Gamma_\ell^R\|^2 = N^{-1}\|\Psi_\ell^R - \Lambda_\ell^R\|^2 \geq C \Big\}. \tag{3.18}$$

This columns of set $\mathcal{Z}$ are essential for the estimation of the structural instability. Since the Step 1.5 of Algorithm 1 makes a transformation of the estimators, we require the following additional assumption:

Assumption 7. *If* $r_a = r_b$, *then* $\inf_{\|W\|=1} N^{-1}\|\Psi^R W - \Lambda_\ell^R\|^2 \geq C$ *for* $\ell \in \mathcal{Z}$. □

Assumption 7 is not restricted by the condition of Λ^R. It is true whenever Λ^R is in the column space generated by Ψ^R or not. It is enforced on the factor loadings Λ^R of the normalized version of the factor model in (3.3) rather than the original factor model in (2.1) and (3.2). Additionally, if no structural instability exists, then $\mathcal{Z}$ is empty and Assumption 7 is not required.

Chapter 4

Proposed Methods in Unknown Break Point Case

In the previous chapter, we consider the case where the break point is known. In this chapter, we generalize the result to deal with the case where the break point is unknown. Similarly, we start with the identification of the instability, the calculation of the shrinkage estimator, the estimators of the number of factors and the break point and PMS estimator. We highlight the different analysis method from the known break point situation. We also point out the distinct implementation of estimating the tuning parameters and two-step estimation algorithm for unknown break point case.

Section 4.1 shows how to identify the instability. In Section 4.2, we present the process of calculating the shrinkage estimator. In Section 4.3, we describe the estimators of the number of factors and the break point. Section 4.4 presents the estimator of

π_0 for PMS. In short, the estimator of π_0 is based on the knowledge of Section 3.4. In Section 4.5, we assess the tuning parameters and clarify some additional procedure of two-step estimation algorithm. Section 4.6 presents some Assumptions and states some theorems which apply the asymptotic theory in case of large sample.

4.1 Identification of the Instability

Given that T is the number of time points in the sample, we introduce a new parameter $\pi = T_0/T \in \Pi$, which stands for the break point fraction, and Π is some closed subset inside [0,1]. Regarding to any $\pi \in \Pi$, we split the full sample by the break point into pre- and post- subsets $X_a(\pi) = (X_1, \cdots, X_{T_a})' \in R^{T_a \times N}$ and $X_b(\pi) = (X_{T_a+1}, \cdots, X_T)' \in R^{T_b \times N}$, where T_a is the integer part of $T\pi$ and $T_b = T - T_a$. To obtain the unknown break point fraction π_0, we need to study the number of factors in the observed $X_a(\pi)$ and $X_b(\pi)$, which are $r_a(\pi)$ and $r_b(\pi)$, with other value of π. $r_a(\pi)$ and $r_b(\pi)$ are defined as the number of non-vanishing eigenvalues of $(NT)^{-1}X_a(\pi)'X_a(\pi)$ and $(NT)^{-1}X_b(\pi)'X_b(\pi)$ as $N, T \to \infty$.

The existing study of Breitung and Eickmeier (2011)[10] claims that the subsample of pre- and post-break observations will have one or more additional factors with a misspecified break point. More specifically, to correctly estimate the break point, we need to minimize the sum of the numbers of pre- and post-break factors by adjusting the potential break point fraction π.

4.2 Shrinkage Estimator

When we estimate the shrinkage estimator, the different situation for unknown break point is that we need to consider an unknown and varying break point fraction $\pi \in \Pi = [\pi_1, \pi_2]$, where $\pi_1 > 0$ and $\pi_2 < 1$.

Suppose that $\widetilde{F}_a(\pi) \in R^{T_a \times k}$ consists of the orthonormalized eigenvectors of $(NT_a)^{-1}X_a(\pi)X_a(\pi)'$ with its first k largest eigenvalues. Similarly, $\widetilde{F}_b(\pi) \in R^{T_b \times k}$ consists of the orthonormalized left eigenvectors of $(NT_b)^{-1}X_b(\pi)X_b(\pi)'$ with its first k largest eigenvalues. The unrestricted estimators of the factor loadings are $\widetilde{\Lambda}_{LS}(\pi) = T_a^{-1}X_a(\pi)'\widetilde{F}_a(\pi)$, $\widetilde{\Psi}_{LS}(\pi) = T_b^{-1}X_b(\pi)'\widetilde{F}_b(\pi)$, and the difference is $\widetilde{\Gamma}_{LS}(\pi) = \widetilde{\Psi}_{LS}(\pi) - \widetilde{\Lambda}_{LS}(\pi)$.

We can apply the procedure in Sections 3.2 with π_0 instead of π. The shrinkage estimator with respect to different $\pi \in \Pi$ estimates the corresponding $r_a(\pi)$ and $r_b(\pi)$. However, this procedure is not ideal because the estimators of $r_a(\pi)$ and $r_b(\pi)$ are highly dependent on π. To overcome this difficulty, we introduce the averaging penalty functions to stabilize the shrinkage estimators. The shrinkage estimators can be obtained by the following function:

$$(\widehat{\Lambda}(\pi), \widehat{\Gamma}(\pi)) = \underset{\Lambda \in R^{N\times k}, \Gamma \in R^{N \times k}}{\operatorname{argmin}} \Big[M(\Lambda, \Gamma; \pi) + P_1^*(\Lambda) + P_2^*(\Gamma) \Big], \tag{4.1}$$

where

$$M(\Lambda, \Gamma; \pi) = (NT)^{-1} \left[\left\| X_a(\pi) - \widetilde{F}_a(\pi)\Lambda' \right\|^2 + \left\| X_b(\pi) - \widetilde{F}_b(\pi)(\Lambda + \Gamma)' \right\| \right], \tag{4.2}$$

$$P_1^*(\Lambda) = \sum_{\ell=1}^{k} \mathbb{E}_\xi \left[\alpha_{NT}(\xi)\omega_\ell^{\lambda *}(\xi)\right] \|\Lambda_\ell\| \text{ and } P_2^*(\Gamma) = \sum_{\ell=1}^{k} \mathbb{E}_\xi \left[\beta_{NT}(\xi)\omega_\ell^{\gamma *}(\xi)\right] \|\Gamma_\ell\|, \quad (4.3)$$

where ξ has uniform distribution on Π and $\mathbb{E}_\xi[\cdot]$ stands for the expected value of the functions of ξ. The terms $P_1^*(\Lambda)$ and $P_2^*(\Gamma)$ represent the averaging penalty functions. The constant parameters $\alpha_{NT}(\pi)$ and $\beta_{NT}(\pi)$ depend on N and T for different value of π. For $\pi \in \Pi$, let $\widetilde{\Lambda}(\pi)$, $\widetilde{\Psi}(\pi)$ and $\widetilde{\Gamma}(\pi)$ be some preliminary estimators. We compute the adaptive weights $\omega_\ell^{\lambda *}(\pi)$ and $\omega_\ell^{\gamma *}(\pi)$ as:

$$\begin{aligned}
\omega_\ell^{\lambda *}(\pi) &= \left(N^{-1}\|\widetilde{\Lambda}_\ell(\pi)\|^2 \mathcal{I}_{\{\widetilde{\Lambda}_\ell(\pi)\neq 0_{N\times 1}\}} + N^{-1}\|\widetilde{\Lambda}_{\ell,LS}(\pi)\|^2 \mathcal{I}_{\{\widetilde{\Lambda}_\ell(\pi)= 0_{N\times 1}\}}\right)^2, \\
\omega_\ell^{\gamma *}(\pi) &= \left(N^{-1}\min\ \left\{\|\widetilde{\Gamma}_\ell(\pi)\|^2, \|\widetilde{\Psi}_\ell(\pi)\|^2\right\}\mathcal{I}_{\{\widetilde{\Gamma}_\ell(\pi)\neq 0_{N\times 1}\}}\right)^{-2} \\
&\quad + \left(N^{-1}\min\ \left\{\|\widetilde{\Gamma}_{\ell,LS}(\pi)\|^2, \|\widetilde{\Psi}_{\ell,LS}(\pi)\|^2\right\}\mathcal{I}_{\{\widetilde{\Gamma}_\ell(\pi)= 0_{N\times 1}\}}\right)^2.
\end{aligned} \quad (4.4)$$

Comparing two versions of the weights in (3.7) and (4.4), it should be noticed that $\omega_\ell^{\lambda *}(\pi_0) = \omega_\ell^{\lambda}$ but $\omega_\ell^{\gamma *}(\pi_0) \neq \omega_\ell^{\gamma}$. When the break point is unknown, we need to use the modified adaptive weights $\omega_\ell^{\gamma *}(\pi)$ for estimation of r_b to generate larger penalties.

4.3 Estimators of the Number of factors and the Break Point

When we analyze the estimators, we consider the influence of varying break point fraction π and adopt *-superscripts to distinguish the one from the known break point case.

The estimators $\widehat{B}^*$, $\widehat{r}_a^*$ and $\widehat{r}_b^*$ can be obtained as follows. The parameter $\mathcal{B}$ is estimated by,

$$\widehat{\mathcal{B}}^* = \mathcal{I}_{\left\{\sup_{\pi\in\Pi}\{\|\widehat{\Gamma}(\pi)\|\}>0\right\}}. \tag{4.5}$$

The number of pre- and post-break factors can be estimated by

$$\widehat{r}_a^* = \min_{\pi\in\Pi} \widehat{r}_a(\pi) \text{ and } \widehat{r}_b^* = \max_{\pi\in\Pi} \widehat{r}_b(\pi), \tag{4.6}$$

where $\widehat{r}_a(\pi)$ and $\widehat{r}_b(\pi)$ are defined as in (3.10), replacing $\widehat{\Lambda}$ and $\widehat{\Gamma}$ by $\widehat{\Lambda}(\pi)$ and $\widehat{\Gamma}(\pi)$ respectively.

4.4 Post Model Selection Estimation

In Section 3.4, we derive the PMS estimator in the known break point situation. If the break point is unknown, the PMS estimator is similar. The only difference is to add the notation *-superscripts and the (π)-arguments.

The PMS estimators are defined as

$$\widehat{\Lambda}_{PMS}(\pi) = \left(\overline{\Lambda}(\pi), \mathbf{0}\right) \text{ and } \widehat{\Psi}_{PMS}(\pi) = \left(\overline{\Psi}(\pi), \mathbf{0}\right), \tag{4.7}$$

where $\mathbf{0}$ is the zero matrix.

Based on the method of Bai (1997)[4], one can estimate π_0 by using the least-squares objective function. Namely, we have

$$\widehat{\pi} = \underset{\pi\in\Pi}{\operatorname{argmin}} Q_{NT}(\pi; \widehat{r}_a^*, \widehat{r}_b^*), \tag{4.8}$$

where

$$
\begin{aligned}
&Q_{NT}(\pi;\widehat{r}_a^*,\widehat{r}_b^*) \\
&\quad = (NT)^{-1}\left[\left\|X_a(\pi)-\widetilde{F}_a(\pi)\widehat{\Lambda}'_{PMS}(\pi)\right\|^2+\left\|X_b(\pi)-\widetilde{F}_b(\pi)\widehat{\Psi}'_{PMS}(\pi)\right\|\right].
\end{aligned} \tag{4.9}
$$

When we are able to detect a break, this procedure can be applied to identify the location of the break point.

4.5 Implementation of the Method and Algorithm

In this section, we generalize the method of estimating the tuning parameters first. Then, we apply the two-step estimation algorithm and calibrate the penalty weights by cross validation. The tuning parameters α_{NT} and β_{NT} dealing with the unknown break point are similar to the known break point case described in Section 3.5. In this case, the tuning parameters are as follows:

$$
\alpha_{NT}(\pi)=\kappa_1(\pi)N^{-1/2}C_{NT_a}^{-3} \text{ and } \beta_{NT}(\pi)=\kappa_2(\pi)N^{-1/2}C_{NT_b}^{-3}, \tag{4.10}
$$

where $\kappa_1(\pi)\in[\underline{\kappa}_1,\overline{\kappa}_1]$ and $\kappa_2(\pi)\in[\underline{\kappa}_2,\overline{\kappa}_2]$ for some $\underline{\kappa}_1,\overline{\kappa}_2<\infty$. The value of $\kappa_1(\pi)$ and $\kappa_2(\pi)$ are given as defined in (3.13) but with $\widetilde{\Lambda}$ and $\widetilde{\Gamma}$ replaced by $\widetilde{\Lambda}(\pi)$ and $\widetilde{\Gamma}(\pi)$.

A two-step procedure in Section 3.6 can be implemented to the model of unknown break point with π_0 replaced by the possible break point fraction π. Next, set the first step preliminary estimators $\widetilde{\Lambda}^{(1)}(\pi)$, $\widetilde{\Psi}^{(1)}(\pi)$ and $\widetilde{\Gamma}^{(1)}(\pi)$ as unrestricted least squares estimators $\widetilde{\Lambda}_{LS}(\pi)$, $\widetilde{\Psi}_{LS}(\pi)$ and $\widetilde{\Gamma}_{LS}(\pi)$; replace ω_ℓ^λ, ω_ℓ^γ, α_{NT} and β_{NT} with $\omega_\ell^{\lambda *}(\pi)$,

$\omega_\ell^{\gamma*}(\pi)$, $\alpha_{NT}(\pi)$ and $\beta_{NT}(\pi)$; replace the PLS criterion function (3.5) with (4.2); and use the numbers of factors $\widehat{r}_a$ and $\widehat{r}_b$ defined in (4.7). Here, the first step numbers of factors $\widehat{r}_a^{(1)}$ and $\widehat{r}_b^{(1)}$ still remain the same with varying value of π according to their definition in (4.6). Hence, we need to attain the first step shrinkage estimators $\widehat{\Lambda}^{(1)}(\pi)$, $\widehat{\Gamma}^{(1)}(\pi)$, $\widehat{r}_a^{(1)}$ and $\widehat{r}_b^{(1)}$. Next, we can access the second-step estimators $\widehat{\Lambda}^{(2)}(\pi)$ and $\widehat{\Gamma}^{(2)}(\pi)$. Finally, we can obtain the selected model with specific estimators $\widehat{\mathcal{B}}^{*}$, $\widehat{r}_a{}^{*}$, and $\widehat{r}_b{}^{*}$ by those two-step PLS estimator $\widehat{\Lambda}^{(2)}(\pi)$ and $\widehat{\Gamma}^{(2)}(\pi)$ following the calculation algorithm in Section 3.5.

Meanwhile, the cross validation procedure is applied. We need to obtain a common value c which works for all possible break dates. For each potential π, the subsample $X_{(-jN)}$ are constructed in a similar way as in Section 4.4, only replacing π_0 by π. For each π and the corresponding value of c, we obtain a selected model. For the cross validation subsample $X_{(jN)}$, the observations that are not inside the conjectured break interval Π are ignored.

We need to estimate the following four parameters before using the shrinkage estimator: the maximum number of potential factors k, the break point fraction interval Π, the domain $\mathcal{C}$ for the tuning constants, and the numbers of sample partitions n_N and n_T. First, k is inclined to be obtained from some preliminary examination of the data. The literature about specifying the value of k is associated with the estimation of the numbers of factors (Stock and Watson, 2012)[31]. Overestimating the value of k results in generating a large number of potential regressors, i.e. the factors, and may weaken the behavior of the shrinkage estimator. If $\widehat{r}_b = k$, then the value of k

may be set too small. Second, the break point fraction interval Π, or in other words the break interval, is determined by the research object. The centre of the interval could be set around year 1984 if we are interested in studying the breaks related to the Great Moderation; or around 2007 if the aim is to analyze the Great Recession. The performance of the estimators can be optimized if we correctly determine a suitable length of the interval. Finally, we choose a set of $\mathcal{C}$, n_N, and n_T that give good result for the Monte Carlo study.

4.6 Asymptotic Theory

In this section, we show that the shrinkage estimator with the averaging penalty in (4.3) can be extended to achieve the estimation of the unknown-break-point model. We use the two-step estimation algorithm detailed in Section 4.5. In Theorem 3.3 for known break point case, we first establish the asymptotic behaviour of the shrinkage estimators $\widehat{\Lambda}(\pi)$ and $\widehat{\Gamma}(\pi)$ for all π. However, for an unknown break date, this step can be skipped. The principal reason is that the estimation of $r_a(\pi)$ and $r_b(\pi)$ can not be accomplished for all π by shrinkage estimators with the averaging penalty terms. The averaging penalties are inclined to be over-penalized if $\pi \neq \pi_0$. However, we can still estimate r_a and r_b because $r_a \leq r_a(\pi)$ and $r_b \leq r_b(\pi)$.

To indicate that the two-step PLS estimator in Section 4.5 generates the estimation of the model, we rewrite Assumption 7 to allow for the impact of the unknown break point and the averaging penalty. That gives us Assumption 8. For any $\pi \in \Pi$, the normalized system in (2.3) can be rewritten as

$$X_a(\pi) = F_a^R(\pi)\Lambda^R(\pi)' + e_a(\pi) \text{ and } X_b(\pi) = F_b^R(\pi)\Psi^R(\pi)' + e_b(\pi), \tag{4.11}$$

where $F_a^R(\pi)$ and $\Lambda^R(\pi)$ are $T_a \times (r_a + r_b)$ and $N \times (r_a + r_b)$ matrices, respectively, and $F_a^R(\pi)$ and $\Psi^R(\pi)$ are $T_a \times (r_a + r_b)$ and $N \times (r_a + r_b)$ matrices, respectively.

Assumption 8. *(i) If $r_a = r_b$, then $\inf_{\pi\in\Pi, \|W\|=1} N^{-1}\|\Psi^R(\pi)W - \Lambda_\ell^R(\pi)\|^2 \geq C$ for $\ell \in \mathcal{Z}$.* $\square$

Assumption 8(i) is a generalization of Assumption 7 by replacing the break point $\pi = \pi_0$ with unknown $\pi \in \Pi$.

We establish a theorem which shows that even for the unknown break point case, we can still get the asymptotic result of the estimators which include the existence of the break point and the numbers of factors. Before stating the theorem, we present Assumption 9 and 10.

Assumption 9. $\mathbb{E}\big[\|F_t^0\|^4\big] \leq C$, $\mathbb{E}\big[\|\overline{F}_t^0\|^4\big] \leq C$ *and there exist nonrandom positive definite matrices Σ_F and $\Sigma_{\overline{F}}$ such that $T^{-1}\sum_{t=1}^{\lfloor T\pi \rfloor} F_t^0 F_t^{0\prime} = \pi\Sigma_F + O_p(T^{-1/2})$ for $\pi \leq \pi_0$ and $T^{-1}\sum_{t=\lfloor T\pi \rfloor+1}^{T} \overline{F}_t^0 \overline{F}_t^{0\prime} = (1-\pi)\Sigma_{\overline{F}} + O_p(T^{-1/2})$ for $\pi \geq \pi_0$, where both $O_p(T^{-1/2})$ terms are uniform over $\pi \in \Pi$.* $\square$

Assumption 10. *Assumption 4 holds with e_a and e_b replaced by $e_a(\pi)$ and $e_b(\pi)$ and Assumption 4(vi) holds uniformly over $\pi \in \Pi$.* $\square$

Theorem 4.1. *Suppose that Assumptions 1, 3, 5, 8, 9 and 10 hold. Then the estimators within the model selected by the two-step estimator in Algorithm 1 satisfy*

$$\lim_{N,T\to\infty} \Pr(\widehat{\mathcal{B}}^* = \mathcal{B}_0) = 1, \lim_{N,T\to\infty} \Pr(\widehat{r_a}^* = r_a) = 1, \lim_{N,T\to\infty} \Pr(\widehat{r_a}^* = r_b) = 1.$$

The proof of this theorem is similar to that given for Theorem 3.4. We also refer to Cheng et al. (2016)[14].

If the difference between π and π_0 is considerable, the averaging penalty terms working in the shrinkage estimators tend to over-penalize the loadings that would be set to zero for $\pi = \pi_0$. It leads to a tendency of underestimating either $r_a(\pi)$ or $r_b(\pi)$ if the conjectured break point is improperly specified. The estimation of the break point can be identified by applying the estimates $\widehat{r_a}^*$ and $\widehat{r_b}^*$ to the least squares objective function in (4.8). In the next chapter, we implement the Monte Carlo simulation to evaluate the finite-sample performance of the shrinkage estimators and analyze the empirical dataset.

Chapter 5

Numerical Results

In the previous chapters, we develop a shrinkage estimator and design estimators of the occurrence of break point and the number of factors. In the first section of Chapter 5, we carry out Monte Carlo Simulations to evaluate the performance of the estimators. Thus, we compute the probability of correctly estimating the numbers of pre- and post- break factors and the existence of the break point. Moreover, we also compute the mean-squared errors for out-of-sample forecasts (MSFE). In the second section, we deal with the real world financial problem, the Great Recession.

5.1 Monte Carlo Simulations

Monte Carlo Simulation is a widely used technique to evaluate the performance of the proposed statistical procedure. The behaviour of the procedure in random samples can be estimated by the empirical process of actually drawing numerous random samples as the experiments (Mahadevan, 1997)[23]. The strategy of doing this is to create an artificial "world," or $pseudo-population$, which simulates the real world, and to conduct numerous trials to investigate how that procedure behaves across samples (Mooney, 1997)[24].

Section 5.1 is divided into two subsections. Section 5.1.1 introduces the design of the factor models and simulations used in the experiments. The simulation outcomes and interpretation are displayed in Section 5.1.2.

5.1.1 Design of the Factor Models

In this section, we describe the design procedure of the factor model and simulations. The factor models of this major paper refer to the paper of Bates et al. (2013)[6], with the modification of stabilizing the structural instability and emphasizing the influence of large breaks instead of small breaks. The factor models have the forms as follows

$$\begin{aligned} \text{Pre-break:}\quad X_{it} = \lambda_i' F_t + e_{it}, \quad F_{t,\ell} = \rho_a F_{t-1,\ell} + u_{t,\ell}, \\ t = 1, \cdots, \lfloor T\pi_0 \rfloor, \quad \ell = 1, \cdots, r_a, \\ \text{Post-break:}\quad X_{it} = \psi_i' \overline{F}_t + e_{it}, \quad \overline{F}_{t,\ell} = \rho_b \overline{F}_{t-1,\ell} + u_{t,\ell}, \\ t = \lfloor T\pi_0 \rfloor + 1, \cdots, T, \quad \ell = 1, \cdots, r_b, \end{aligned} \tag{5.1}$$

where $i = 1, \cdots, N$, $F_t = (F_{t,1}, \cdots, F_{t,r_a})'$, $\overline{F}_t = (\overline{F}_{t,1}, \cdots, \overline{F}_{t,r_b})'$. For $t = 1, \cdots, T$ and $\ell = 1, \cdots, r_b$, $u_{t,\ell}$ are i.i.d as $N(0, 1)$. To take into account for the time dimensional and cross-sectional dependence of the errors, we suppose that

$$e_{it} = \alpha e_{it-1} + v_{it},\ v_{it} = (v_{1t}, \cdots, v_{Nt})' \sim N(0, \Omega), \tag{5.2}$$

where the (i, j)-th element of Ω is $\beta^{|i-j|}$. We suppose that the processes $\{u_{t,\ell} : \ell = 1, \cdots, r_b\}$ and $\{v_{it} : i = 1, \cdots, N\}$ in (5.2) are mutually independent and are i.i.d across t. F_0 and $e_0 = (e_{10}, \cdots, e_{N0})'$ denote the initial values of the factors and the errors. We suppose that they are drawn from the stationary distribution. If there is no break, which means $r_b = r_a$, then $\overline{F}_{T_0} = F_{T_0}$. If there are some new factors after the break point, which means $r_b > r_a$, then $\overline{F}_{T_0} = (F'_{T_0}, F^{*'}_{T_0})'$, where each element of $F^*_{T_0}$ is drawn independently from the distribution of $F_{t,\ell}$.

First, we explain how to construct the pre-break factor loadings $\{\lambda_i : i = 1, \cdots, N\}$. Let $\lambda_i \sim N(0, \Sigma_i)$ independently, where Σ_i is a diagonal matrix with distinctive elements $\sigma_i^2(1), \cdots, \sigma_i^2(r_a)$. The sum of these diagonal elements controls the population regression R_i^2 of X_{it} on the factors. Here, R_i^2 is the pre-specified regression R^2 of the i-th series. The method of selecting R_i^2 for $i = 1, \cdots, N$ is drawn by Bai and Ng (2002)[5] that R_i^2 is set to 0.5.

Second, we demonstrate the construction of post-break factor loadings ψ_i, which requires to consider the existence of the structural instability. We set $\psi_i = (1 - w)\lambda_i + w\lambda_i^*$, where λ_i^* has the same distribution as λ_i but they are independent. We can control the size of the instability by changing the value of the scalar w. Particularly, $w = 0$ corresponds to no break in the factor loadings.

The simulated time series has mean 0 and variance 1. We extract a maximum of $k = 8$ potential factors from the sample. For known break point, the estimators is based on the two-step PLS estimator described in Algorithm 1. We set $n_N = 5$ and $n_T = 10$, which means we have five cross-sectional partitions and ten time dimensional partitions. Normally, considering the cross-sectional dimension, the procedure has to be implemented on each partition of the sample, so it is time consuming. On the contrary, the analysis based on time partitions is fast.

For the unknown break point case, the estimation depends on adjusted version of Algorithms 1 described in Section 4.4. We consider a discrete set Π_d which has the grid size $\tau = 0.001$. It is a shift by a quarter for a monthly data set of 300 periods. We set $\Pi_d = \{\pi_0 - 4\tau, \pi_0 - 3\tau, \cdots, \pi_0, \cdots, \pi_0 + 3\tau, \pi_0 + 4\tau\}$, which is symmetric around the true break point π_0 and it stretches over a two-year interval in total. Moreover, the post-break subsample of the PMS estimator is attained by the least square estimator of the break point described in Section 4.3.

To evaluate the performance of the selected model, we also compute the mean-squared errors for out-of-sample forecasts (MSFE). The forecast series follows

$$\begin{aligned} \text{Initial Value:} \quad & y_1 = X_{iT} \\ \text{Pre-break:} \quad & y_{t+1} = \varphi_a' F_t + \epsilon_{t+1}, \quad t = 1, \cdots, T_a, \\ \text{Post-break:} \quad & y_{t+1} = \varphi_b' \overline{F}_t + \epsilon_{t+1}, \quad t = T_a + 1, \cdots, T_a + T_b. \end{aligned} \tag{5.3}$$

Suppose that the errors ϵ_1, $\epsilon_2, \cdots, \epsilon_{T_a+T_b}$ are iid as $N(0,1)$. The loading vector is generated from the distribution $\varphi_a \sim N(0, I_{r_a})$. If there is no structural break, then we have unchanging factor loadings, which means $\varphi_b = \varphi_a$. If we consider the structural instability defined in (2.3), then the post-break factor loading $\varphi_b = (1-w)\varphi_a + w\varphi_a^*$, where φ_a^* and φ_a are independent and have the same distribution.

To evaluate the MSFE of the out-of-sample forecasts, first, we select the model and the factors based on the X sample. Second, if there is no break points, we estimate $\varphi_b = \varphi_a$ based on the full sample $t = 1, \cdots, T_a + T_b - 1$ and evaluate the MSFE associated with the prediction of $y_{T_a+T_b+1}$. If there exists a break point, we estimate φ_b based on the subsample $t = T_a + 1, \cdots, T_a + T_b - 1$ and evaluate the MSFE associated with the prediction of $y_{T_a+T_b+1}$.

$$\mathrm{MSFE}(\widehat{y}_{T_a+T_b+1}) = \mathbb{E}\big[(y_{T_a+T_b+1} - \widehat{y}_{T_a+T_b+1})^2\big]. \tag{5.4}$$

In the simulation, we report the relative MSFE in (5.5), which is based on both the PMS estimator and the full-sample estimation. The full-sample estimator $\widehat{\Lambda}$ is defined as the first r columns of the full sample least squares estimator $\widetilde{\Lambda}_{LS} = T^{-1}X'\widetilde{F}$, where $r = r_a$ if $\mathcal{B}_0 = 0$ (no break), $r = r_a + r_b$ otherwise. If the value of the relative MSFE is less than 1, then the predictor based on the PMS estimator domains the full sample estimator.

$$\text{Relative MSFE} = \mathrm{MSFE}_{PMS}/\mathrm{MSFE}_{full} \tag{5.5}$$

5.1.2 Results for Shrinkage Estimator

In this section, we present the results of three types of Monte Carlo experiments listed in Table 5.1. In the first experiment, we assume the break point is known as located in the middle of the sample ($\pi = 0.5$). The cross-sectional correlation is set to be modest that α and β equal to 0.2. The second experiment has the same rate of correlation. In particular, we first consider that the break point is known and located nearly at the end of the sample ($\pi_0 = 0.8$). Second, we consider the scenario where the break point is unknown. Finally, besides the choice of a stronger cross-sectional correlation ($\alpha, \beta = 0.5$), the third experiment is similar to the first one. We suppose that the temporal correlation ρ_a and ρ_b are equal to 0.5 in all three cases, and we implement 1,000 Monte Carlo experiments to take the average result.

Table 5.1: Monte Carlo Experiments

Exp.	π_0	α, β	Break Point
1	0.5	0.2	Known
2	0.8	0.2	Known, Unknown
3	0.5	0.5	Known

Notes: π_0: break point fraction; α, β: cross-sectional correlation; temporal correlation $\rho_a = \rho_b = 0.5$.

First, we display the Monte Carlo results of Experiment 1 in Table 5.2. The Panel A and B correspond to the situations of no break and the existence of instability respectively. We set the scalar of the instability w as 0.5 and 0.7 and we include various choices of N and T in the experiment to test the asymptotic behaviour. We report the probability of correctly determining $\mathcal{B}_0$, r_a and r_b. The last column gives the MSFE of the predictor based on the PMS estimator relative to the predictor based on a forecast of full-sample least-squares estimator, where the number of factors is set to r_a for Panel A, and to $r_a + r_b$ for Panel B. The values less than 1 mean that the proposed PMS predictor is more accurate.

Table 5.2: Known Break Point, $\pi_0 = 0.5$

Model Configuration					Model Selection			Relative
r_a	r_b	w	N	T	$\Pr(\widehat{\mathcal{B}} = \mathcal{B}_0)$	$\Pr(\widehat{r_a} = r_a)$	$\Pr(\widehat{r_b} = r_b)$	MSFE
Panel A. No Break								
3	3	0	120	120	1.00	1.00	1.00	1.00
3	3	0	140	140	1.00	1.00	1.00	1.00
3	3	0	180	180	1.00	1.00	1.00	1.00
Panel B. Instability								
3	3	0.5	120	120	1.00	1.00	1.00	0.38
3	3	0.5	140	140	1.00	1.00	1.00	0.98
3	3	0.5	180	180	1.00	1.00	1.00	1.00
3	3	0.7	120	120	1.00	1.00	1.00	0.39
3	3	0.7	140	140	1.00	1.00	1.00	0.80
3	3	0.7	180	180	1.00	1.00	1.00	1.06

Notes: Cross-sectional correlation $\alpha = \beta = 0.2$; temporal correlation $\rho_a = \rho_b = 0.5$.

The procedure is very accurate in estimating the estimators $\widehat{\mathcal{B}}$, $\widehat{r_a}$ and $\widehat{r_b}$ in all cases with chosen value of w. The possible reason is that the selection procedure has strong power which can deal with big changes in the loadings. When the break in the factor loadings is rather big ($w = 0.5$ or 0.7), it identifies the break point and the number of factors with probability 1. In terms of the relative MSFEs, it should be noticed that the proposed PMS predictor are almost accurate since the relative MSFEs is approximately equal to or less than 1. It means that the PMS predictor dominates the full-sample predictor.

Second, we move to the known break point case of Experiment 2. Table 5.3 gives the probability of $\widehat{\mathcal{B}} = \mathcal{B}_0$, $\widehat{r_a} = r_a$ and $\widehat{r_b} = r_b$ with instability scalar 0 (no break), 0.7 and 1, as well as the MSFE.

Table 5.3: Known Break Point, $\pi_0 = 0.8$

Model Configuration					Model Selection			Relative
r_a	r_b	w	N	T	$\Pr(\widehat{\mathcal{B}} = \mathcal{B}_0)$	$\Pr(\widehat{r}_a = r_a)$	$\Pr(\widehat{r}_b = r_b)$	MSFE
Panel A. No Break								
3	3	0	120	240	1.00	1.00	1.00	1.00
3	3	0	120	300	1.00	1.00	1.00	1.00
3	3	0	150	300	1.00	1.00	1.00	1.00
Panel B. Instability								
3	3	0.7	120	240	0.00	1.00	1.00	1.05
3	3	0.7	120	300	1.00	1.00	1.00	0.69
3	3	0.7	150	300	1.00	1.00	1.00	1.39
3	3	1	120	240	1.00	1.00	1.00	0.55
3	3	1	120	300	1.00	1.00	1.00	0.36
3	3	1	150	300	1.00	1.00	1.00	0.61

Notes: Cross-sectional correlation $\alpha = \beta = 0.2$; temporal correlation $\rho_a = \rho_b = 0.5$.

When there is no break point, experiment 2 turns out that the procedure is overall accurate to determine $\mathcal{B}_0$, r_a and r_b. With a variety value of N and T, the estimators of B, r_a and r_b are almost accurate with probability 1. Moreover, the PMS estimator weakly dominates the full-sample predictor since the value of MSFE are slightly less than 1. Compared with the instability scale $w = 0.7$, when it is set to be 1, the proposed PMS predictor does better than the full-sample least squares estimator. We can tell that a larger value of w is necessary for the estimation to be accurate.

Table 5.4 is the output of Experiment 2 which shows that an unknown break point leads the result to be less accurate. Under the no-change scenario, the procedure correctly determines the estimators for all three sample sizes. However, to take into account the structural instability, it seems that a larger break in the loadings ($w = 1$ instead of $w = 0.7$) and larger sample size ($N > 120$ and $T > 240$) is preferred to increase the probability of detecting the break. Besides that, the result of estimating the numbers of factors for all cases is perfect. According to the last column of Table 5.4, the PMS predictor performs well under the no-break scenario since the MSFEs equal to 1, which means that the PMS predictor is equivalent to the full-sample predictor. When we consider the instability, for both cases $w = 0.7$ and $w = 1$, the proposed PMS estimator dominates the full-sample predictor.

Table 5.4: Unknown Break Point, $\pi_0 = 0.8$

Model Configuration					Model Selection			Relative
r_a	r_b	w	N	T	$\Pr(\widehat{\mathcal{B}} = \mathcal{B}_0)$	$\Pr(\widehat{r}_a = r_a)$	$\Pr(\widehat{r}_b = r_b)$	MSFE
Panel A. No Break								
3	3	0	120	240	1.00	1.00	1.00	1.00
3	3	0	120	300	1.00	1.00	1.00	1.00
3	3	0	150	300	1.00	1.00	1.00	1.00
Panel B. Instability								
3	3	0.7	120	240	0.00	1.00	1.00	0.79
3	3	0.7	120	300	0.00	1.00	1.00	0.89
3	3	0.7	150	300	0.00	1.00	1.00	0.59
3	3	1	120	240	0.50	1.00	1.00	1.17
3	3	1	120	300	1.00	1.00	1.00	0.42
3	3	1	150	300	1.00	1.00	1.00	0.18

Notes: Cross-sectional correlation $\alpha = \beta = 0.2$; temporal correlation $\rho_a = \rho_b = 0.5$.

Finally, we present the result of experiment 3, which is very similar to experiment 1 (Table 5.1) but only with stronger dependence and correlation in errors ($\alpha = \beta = 0.5$). The results are reported in Table 5.5. In general, the procedure correctly estimates $\mathcal{B}_0$, r_a and r_b. All the value of MSFEs are less than 1, so the results favor the proposed PMS predictor. The noticeable thing is that compared to Table 5.1, we obtain less accurate detection of the break point and the estimation of the number of post-break factors in the case of $N, T = 140$ due to the increased cross-sectional correlation.

Table 5.5: Known Break Point, $\pi_0 = 0.5$

Model Configuration					Model Selection			Relative
r_a	r_b	w	N	T	$\Pr(\widehat{\mathcal{B}} = \mathcal{B}_0)$	$\Pr(\widehat{r}_a = r_a)$	$\Pr(\widehat{r}_b = r_b)$	MSFE
Panel A. No Break								
3	3	0	120	120	1.00	1.00	1.00	1.00
3	3	0	140	140	1.00	1.00	1.00	1.00
3	3	0	180	180	1.00	1.00	1.00	1.00
Panel B. Instability								
3	3	0.5	120	120	1.00	1.00	1.00	0.10
3	3	0.5	140	140	0.50	1.00	0.50	0.79
3	3	0.5	180	180	1.00	1.00	1.00	0.21
3	3	0.7	120	120	1.00	1.00	1.00	0.16
3	3	0.7	140	140	1.00	1.00	1.00	0.95
3	3	0.7	180	180	1.00	1.00	1.00	0.28

Notes: Cross-sectional correlation $\alpha = \beta = 0.5$; temporal correlation $\rho_a = \rho_b = 0.5$.

5.2 Real Data Set: the Great Recession

The Great Recession in the United States started in December 2007 and finished in June 2009. It origins in an unusually dramatic financial crisis, which began with a financial collapse that erased more that half the capitalization of the stock market (Grusky et al. 2011)[19]. It is known as the longest post-war recession and its negative impact to the labor market is huge. From May 2007 to October 2009, the labor force lost 7.5 million jobs and more, and the unemployment rate ascended from 4.4% to 10.1% (Roberts 2009)[27]. From the global view, the world economy declined by 6% (Roberts 2009)[27]. We pick the Great Recession as an example for the empirical analysis because of its profound and large-scale influence.

Based on the data set of the Great Recession, we use the procedures developed in Chapters 2 to 4 to investigate the fluctuation of factor loadings and the emergence of new factors. Section 5.2.1 gives some preliminary transformations and the analysis results are given in Section 5.2.2.

5.2.1 Some Preliminary Transformations

The data set of the Great Recession is from Stock and Watson (2012)[31], which contains a set of 200 macroeconomic and financial indicators. As in Cheng et al. (2016)[14]. After eliminating the duplicate series, we obtain the data set SW132 which includes 132 out of 200 indicators. We extend SW132 to 2012:M12, which represents the twelfth month of year 2012, using May 2013 data vintages. For consumption of durables, nondurables, services, nonresidential investment and 16 price series, we replace the quarterly series by their monthly counterparts. The rest quarterly series, which are not able to be converted to monthly series, are removed. Then, we include two new monthly components: change in private inventory and wage and salary disbursements. Following Stock and Watson (2012)[31], we remove local means from all series using a biweight kernel with a bandwidth of 100 months. The local means are approximately the same as the ones obtained by a centered moving average of ± 70 months. The data set consists the observations of $N = 102$ series of monthly financial indicators, e.g. growth rates of participation of recreation services or monthly inflation of sales of retail sotres. The sample begins after the Great Moderation and ranges from January 1985 to January 2013 which lasts 337 months ($T = 337$). A detailed list of the financial indicators is presented in the appendix.

5.2.2 Analysis Results

The two step estimation procedure described in Section 3.5 is applied. Moreover, when considering unknown break point, we also combine the procedure summarized in Section 4.4. We fix the number of potential factors $k = 8$ and use $n_N = 5$ and $n_T = 10$. The model selection results are presented in Table 5.6.

Table 5.6: Model Selection, T_c is 2007:M12

Interval (Month)		Factors		Break Points	
Size	Range	$\widehat{r}_a$	$(\widehat{r}_b - \widehat{r}_a)$	Least Sq.$(\widehat{\pi})$	Revised
0	2007:M12-2007:M12	1	1	2007:M12	2007:M12
3	2007:M9-2008:M3	1	1	2007:M9	2007:M12
6	2007:M6-2008:M6	1	1	2007:M6	2007:M12
9	2007:M3-2008:M9	1	2	2007:M3	2007:M12

Notes: We center the interval at 2007:M12 and use the averaging penalty functions $P_1^*(\Lambda)$ and $P_2^*(\Lambda)$ defined in (4.3) where the average is taken over the interval 2007:M12 $\pm$ Size.

According to the U.S. National Bureau of Economic Research (the official arbiter of U.S. recessions), December 2007 is the beginning of the Great Recession. Therefore, the picked interval of potential break points are centered around $T_c = $ 2007:M12 with four different interval size. For the first case in Table 5.6, the interval contains a single month (2007:M12), meaning that we consider the break point happens in 2007:M12. For interval size 9, the potential break point located in the months from 2007:M3 to 2008:M9. From the result, we attain one pre-break factor ($\widehat{r}_a = 1$) for all variety of interval sizes, and two or three post-break factors ($\widehat{r}_b = 2$ or $\widehat{r}_b = 3$). There is a strong evidence of a change in the number of factors. In columns 4 and 5 of Table 5.6, we report the least squares estimation of the break point $\widehat{\pi}$ defined in (4.8). In the fifth column, we minimize the least square criterion over the chosen interval.

We can tell from Table 5.6 that we obtain less accurate result in estimating the correct break points by using least square method when we have intervals with nonzero size. It turns out that the minimum is always attained at the boundary. Moreover, we revise the break point by using the properties of $\widehat{r}_a$ and $\widehat{r}_b$. In Section 4.1, we derive from the study of Breitung and Eickmeier (2011)[10] that the sum of pre- and post-break factors k is minimized at the true break point. Thus, for each given interval, we compute $\widehat{r}_a + \widehat{r}_b$ for each potential break point. Then we check whether the minimum of the sum over this interval is obtained at T_c = 2007:M12. If so, we set the revised break point equal to T_c. Otherwise, we adjust the revised break point as the closest point to T_c where the minimum is attained. For all four cases in the analysis, the break points we obtained match the conjectured break point T_c (Table 5.6). Therefore, the procedure has strong capability of locating unknown break point and estimating the number of factors.

Next, we applied bootstrap method based on the original dataset SW132. According to Efron and Tibshirani (1993)[16], Bootstrap is a recently developed simulation methodology using computer power to realize statistical inferences and assign measures of accuracy to statistical estimates. The process of bootstrap includes two steps. First, we draw a bootstrap sample, which has the same size as the initial observations, with replacement. Then, we repeat this process a large number of times to get bootstrap sample replicates which can be used to make inferences (Efron and Tibshirani, 1993)[16]. This technique requires fewer assumptions and offer greater accuracy and insight than standard methods in many fields (Stine, 1989)[30].

Concretely, we first generate 10000 bootstrap samples, then obtain the number of factors $\widehat{r}_a$, $\widehat{r}_b$, and the occurrence of break point $\widehat{\mathcal{B}}$ from each sample by applying the same procedure. We only focus on the situations that the break point is unknown. To calculate the probability of correctly estimating the number of factors, we round off the mean of $\widehat{r}_a$ and $\widehat{r}_a$ and set those value as r_a and r_b respectively. The value of MSFE shows the behaviour of the procedure. The result is displayed in Table 5.7 below.

Table 5.7: Bootstrap Results

Model Configuration				Model Selection			Relative
r_a	r_b	N	T	$\Pr(\widehat{\mathcal{B}} = \mathcal{B}_0)$	$\Pr(\widehat{r}_a = r_a)$	$\Pr(\widehat{r}_b = r_b)$	MSFE
Panel A. Unknown Break							
3	3	102	337	0.85	0.41	0.52	0.85

From Table 5.7, it is obvious that we tend to properly detect the break point as the rate of success is higher that 85% for unknown break point. However, we have relatively lower probability of accessing the number of pre- and post- break factors which is around 0.4-0.5. In terms of the MSFE, the value is 0.85 which is less that 1. Therefore, for the bootstrap samples, the proposed PMS estimators domain the corresponding full sample estimators.

Chapter 6

Conclusion

In case of macroeconomics, one may be interested in studying the occurrence of the structural break, such as the 2007-2009 Great Recession, and its influence to the financial market and the whole economic environment. In order to eliminate the negative effect of the break point and maintain the economic stability, one needs to keep track of any new factors and the change of existing factors after the break point.

A high-dimensional model system is adopted to solve this problem. A shrinkage estimation procedure identifies the occurrence of the break point and estimates the numbers of pre- and post-break factors. The estimator works well as it is weakly dependent on the instabilities of unknown break point case. When the number of factors remains unchanged, the procedure can detect the changes in the factor loadings. We also find out that the occurrence of the break point can be identified by using a conventional least squares approach once we obtain the estimation of the numbers of pre- and post-break factors.

We applied Monte Carlo Simulation to analyze the behaviour of the procedure. It turns out that the shrinkage procedure has capability of determining the number of factors and testing for a break point for both known- and unknown-break-point case. In the empirical analysis based on the 2007-2009 Great Recession data set, the procedure can properly detect the break point in the most of the cases.

Appendix A

Some Statistical Background

In this appendix, we give some definitions, lemmas and theorems used in deriving the main results of this major paper. Most of the definitions, can be found in statistical textbooks, such as

Definition A.1 (Casella and Berger (2002)[12]). *A sequence of random variables* $\{X_n\}_{n=1}^{\infty}$ *converges in probability to a random variable* X *if, for every* $\epsilon > 0$,

$$\lim_{n\to\infty} \mathrm{P}(|X_n - X| \geq \epsilon) = 0.$$

We denote it as $X_n \xrightarrow[n\to\infty]{\mathrm{p}} X$.

Definition A.2 (Bickel and Doksum (2001)[8]). *A sequence of random vectors* $Z_n = (Z_{n1}, Z_{n2}, \cdots, Z_{nm})'$ *converges in probability to* $Z = (Z_1, Z_2, \cdots, Z_m)'$ *iff* $Z_{nj} \xrightarrow[n\to\infty]{\mathrm{p}} Z_j$ *for* $1 \leq j \leq m$. *We denote it as* $Z_n \xrightarrow[n\to\infty]{\mathrm{p}} Z$.

Definition A.3 (Fuller (1976)[18]). *Let $\{A_n\}_{n=1}^{\infty}$ be a sequence of random matrices and $\{b_n\}_{n=1}^{\infty}$ be a sequence of positive real numbers.*

(i) $A_n = \mathrm{O}_p(b_n)$ if for all i and j $(A_n)_{ij} = \mathrm{O}_p(b_n)$, that is, $\forall \epsilon > 0$, there exists $k_\epsilon > 0$ and an integer N_ϵ, such that $\mathrm{P}\big(|\,(A_n)_{ij}| > k_\epsilon b_n\big) < \epsilon$, $\forall n > N_\epsilon$.

(ii) $A_n = \mathrm{o}_p(b_n)$ if for all i and j $(A_n)_{ij} = \mathrm{o}_p(b_n)$, that is, $\frac{(A_n)_{ij}}{b_n} \xrightarrow[n\to\infty]{\mathrm{p}} 0$.

Lemma A.1 (Strawderman (1994)[32]). *Let A_n be a random sequence of symmetric nonnegative definite $k \times k$ matrices where $k < \infty$. If a positive definite symmetric $k \times k$ matrix A with finite elements exists such that $A_n \xrightarrow{P} A$ element-wise, then $\|A_n - A\| \xrightarrow{P} 0$, where $\|\cdot\|$ denotes any proper norm on $\mathbb{R}^{k\times k}$.*

The proof of this lemma is given as Lemma 1 in Strawderman (1994)[32].

Appendix B

Some Proofs

In this appendix, we give the proof of one of the theorems used in the major paper.

Proof of Theorem 3.4. We need to consider two situations: there exists a break point with the instability of change and there is no change. In each situation, we have to prove that $\lim_{N,T\to\infty} \Pr(\widehat{r_a} = r_a) = 1$, $\lim_{N,T\to\infty} \Pr(\widehat{r_b} = r_b) = 1$ and $\lim_{N,T\to\infty} \Pr(\widehat{\mathcal{B}} = \mathcal{B}_0) = 1$.

First, we consider the change with instability. Let $\Sigma_a = \Lambda^{0\prime}\Lambda^0/N \in R^{r_a \times r_a}$, where $\Sigma_a^{1/2}$ is the Cholesky factor of Σ_a. Let $\Upsilon_a = [v_1\ v_2\ \cdots\ v_{r_a}]$ be a matrix of orthonormal eigenvectors such that ${\Upsilon_a}'(\Sigma_a^{1/2})'\Sigma_F\Sigma_a^{1/2}\Upsilon_a = V_a$. Suppose we have the eigenvalues of $(\Sigma_a^{1/2})'\Sigma_F\Sigma_a^{1/2}$ ordered from largest to smallest denoted by ρ_ℓ, $\ell = 1, \cdots, r_a$. Therefore, by definition, V_a is a diagonal matrix of eigenvalues of $(\Sigma_a^{1/2})'\Sigma_F\Sigma_a^{1/2}$ ordered from largest to smallest.

Further, one can verify that ρ_ℓ, the ℓ-th largest eigenvalue of $(\Sigma_a^{1/2})'\Sigma_F\Sigma_a^{1/2}$, is also the

ℓ-th largest eigenvalue of matrix $\Sigma_a \Sigma_F$. Then, we define the transformation matrix R_a and R_b in (3.3) such that $R_a = \Sigma_a^{1/2} \Upsilon_a V_a^{-1/2}$ and $R_b = \Sigma_b^{1/2} \Upsilon_b V_b^{-1/2}$. Moreover, we have defined that $\Lambda^R = \Lambda^0 (R_a^{-1})' \in R^{N \times r_a}$ and $\Psi^R = \Psi^0 (R_b^{-1})' \in R^{N \times r_b}$ in (3.3) and (3.9).

Then, we have

$$\begin{aligned} \frac{\Lambda^{R\prime} \Lambda^R}{N} &= \frac{R_a^{-1} \Lambda^{0\prime} \Lambda^0 (R_a^{-1})'}{N} = V_a^{1/2} \Upsilon_a' \Sigma_a^{-1/2} \frac{\Lambda^{0\prime} \Lambda^0}{N} \Sigma_a^{-1/2} \Upsilon_a V_a^{1/2} \\ &= V_a^{1/2} \Upsilon_a' \Sigma_a^{-1/2} \Sigma_a \Sigma_a^{-1/2} \Upsilon_a V_a^{1/2} = V_a. \end{aligned} \tag{B.1}$$

By Assumption 3, we have $\|\Sigma_a - \Sigma_\Lambda\| \to 0$ as $N \to \infty$. Since Σ_a is a sequence of symmetric positive definite matrices, we can apply Lemma 1 from Strawderman (1993)[32] to obtain $\Sigma_a \xrightarrow[N\to\infty]{P} \Sigma_\Lambda$. Then we multiply the positive definite matrix Σ_F on both sides to get $\Sigma_a \Sigma_F \xrightarrow[N\to\infty]{P} \Sigma_\Lambda \Sigma_F$. Since the convergence of matrices is entry-wise (de Boor, 2002)[9], we have for all $\epsilon > 0$,

$$\lim_{N,T\to\infty} \Pr\Big(\big| (\Sigma_a \Sigma_F)_{ij} - (\Sigma_\Lambda \Sigma_F)_{ij} \big| \le \epsilon \Big) = 1. \tag{B.2}$$

where the ij-subscript represents the entry of the matrix located in i-th row and j-th column.

We apply the property that the eigenvalues of a matrix depend continuously to its entries (Alexanderian 2013)[2], we get

$$\rho_\ell(\Sigma_a \Sigma_F) \xrightarrow[N\to\infty]{P} \rho_\ell(\Sigma_\Lambda \Sigma_F). \tag{B.3}$$

Therefore, each eigenvalue of $\Sigma_a \Sigma_F$ ordered from largest to smallest, which is also the ℓ-th diagonal elements of V_a, converges in probability to the ℓ-th largest eigenvalue of $\Sigma_\Lambda \Sigma_F$ as $N \to \infty$, denoted by $\rho_\ell(\Sigma_\Lambda \Sigma_F)$.

Let a_ℓ be a selection vector that selects the ℓ-th column of a matrix. Similarly, a_ℓ' selects the ℓ-th row of a matrix. Then, for $\ell = 1, \cdots, r_a$,

$$\begin{aligned} N^{-1}\|\Lambda_\ell^R\|^2 &= N^{-1}(\Lambda_\ell^{R\prime}\Lambda_\ell^R) = N^{-1}(a_\ell'\Lambda^{R\prime}\Lambda^R a_\ell) = a_\ell'(N^{-1}\Lambda^{R\prime}\Lambda^R)a_\ell = a_\ell' V_a a_\ell \\ &= \rho_\ell(\Sigma_\Lambda\Sigma_F) + o(1). \end{aligned} \tag{B.4}$$

Here, $a_\ell' V_a a_\ell$ means the ℓ-th diagonal element of V_a, which equals to the ℓ-th largest eigenvalue of $\Sigma_\Lambda \Sigma_F$. Equation (B.4) is equivalent to

$$N^{-1/2}\|\Lambda_\ell^R\| = \left[\rho_\ell(\Sigma_\Lambda\Sigma_F)\right]^{1/2} + o(1). \tag{B.5}$$

From Theorem 3.3(i), we obtain

$$N^{-1/2}\|\widehat{\Lambda}_\ell - \Lambda_\ell^R\| = O_p(C_{NT}^{-1}) \text{ for } \ell = r_a. \tag{B.6}$$

By applying the triangle inequality on (B.6), we can get

$$\left| N^{-1/2}\|\widehat{\Lambda}_\ell\| - N^{-1/2}\|\Lambda_\ell^R\| \right| \le N^{-1/2}\|\widehat{\Lambda}_\ell - \Lambda_\ell^R\| = O_p(C_{NT}^{-1}). \tag{B.7}$$

We get rid of the absolute value and only retain the left side. Then, based on (B.5), we obtain the following result

$$\begin{aligned} -N^{-1/2}\|\widehat{\Lambda}_\ell - \Lambda_\ell^R\| &\leq N^{-1/2}\|\widehat{\Lambda}_\ell\| - N^{-1/2}\|\Lambda_\ell^R\| \\ N^{-1/2}\|\Lambda_\ell^R\| - N^{-1/2}\|\widehat{\Lambda}_\ell - \Lambda_\ell^R\| &\leq N^{-1/2}\|\widehat{\Lambda}_\ell\| \\ \left[\rho_\ell(\Sigma_\Lambda \Sigma_F)\right]^{1/2} + o(1) &\leq N^{-1/2}\|\widehat{\Lambda}_\ell\| + O_p(C_{NT}^{-1}). \end{aligned} \tag{B.8}$$

Therefore, we conclude that for $\ell = r_a$, $\lim\limits_{N,T\to\infty} \Pr(\|\widehat{\Lambda}_\ell\| > 0) = 1$. Here, it shows that the r_a-th column of $\widehat{\Lambda}$ has value greater than 0 with probability approaching to 1. Recall the definition from (3.10), $\widehat{r}_a$ is the largest column number of $\widehat{\Lambda}$ which has value great than 0. Thus,

$$\lim_{N,T\to\infty} \Pr(\widehat{r}_a \geq r_a) = 1. \tag{B.9}$$

From Theorem 3.3(ii), we know that

$$\lim_{N,T\to\infty} \Pr(\|\widehat{\Lambda}_\ell\|^2 = 0) = 1 \text{ for } \ell = r_a + 1, \cdots, k. \tag{B.10}$$

Hence, we have the knowledge that the columns of $\widehat{\Lambda}$ from $r_a + 1$ to k has value 0 with probability approaching to 1. In this case,

$$\lim_{N,T\to\infty} \Pr(\widehat{r}_a \leq r_a) = 1. \tag{B.11}$$

By (B.9) and (B.11), we obtain

$$\lim_{N,T\to\infty} \Pr(\widehat{r}_a = r_a) = 1. \tag{B.12}$$

Note that by definition in (3.10), we have $\widehat{r}_b \geq \widehat{r}_a$. Together with (B.12) and the condition $r_a = r_b$, we have

$$\lim_{N,T\to\infty} \Pr(\widehat{r}_b \geq r_b) = 1. \tag{B.13}$$

Similar as (B.11), Theorem 3.3(v) implies that

$$\lim_{N,T\to\infty} \Pr(\widehat{r}_b \leq r_b) = 1). \tag{B.14}$$

From (B.13) and (B.13) we conclude that

$$\lim_{N,T\to\infty} \Pr(\widehat{r}_b = r_b) = 1. \tag{B.15}$$

For the second situation, with the existence of the instability, we have $r_a = r_b$ and $\mathcal{B}_0 = 1$. First note that if $r_a = r_b$, we have

$$N^{-1}\Gamma^{R\prime}\Gamma^R = N^{-1}(\Psi^R - \Lambda^R)'(\Psi^R - \Lambda^R) = \mathbf{e}'\Sigma^+_{\Lambda\Psi}\mathbf{e} + o(1), \tag{B.16}$$

where $\mathbf{e} = \lim_{N\to\infty}(R_a^{-1}, -R_b^{-1})'$ has full rank following Assumption 2 and 3. $\Sigma^+_{\Lambda\Psi}$ is defined in (3.1). By a Cholesky decomposition, write $\Sigma^+_{\Lambda\Psi} = (\Sigma^+_{\Lambda\Psi})^{1/2}(\Sigma^+_{\Lambda\Psi})^{1/2}$ with $\text{rank}((\Sigma^+_{\Lambda\Psi})^{1/2}) = \text{rank}(\Sigma^+_{\Lambda\Psi}) > r_a$. For a $2r_a \times 2r_a$ matrix $(\Sigma^+_{\Lambda\Psi})^{1/2}$, the rank of the null space of $(\Sigma^+_{\Lambda\Psi})^{1/2}$ is smaller than r_a. It follows that $(\Sigma^+_{\Lambda\Psi})^{1/2}\mathbf{e} \neq 0$ because $\text{rank}(\mathbf{e}) = r_a$, and this immediately implies that with $\Sigma_\Gamma = \mathbf{e}'\Sigma^+_{\Lambda\Psi}\mathbf{e} \neq 0$, $r_a > r_b$ and $\text{rank}(\Sigma_{\Lambda\Psi^+}) > r_a$, then

$$N^{-1}\Gamma^{R\prime}\Gamma^R \to \Sigma_\Gamma \text{ for some } \Sigma_\Gamma \neq 0 \text{ as } N \to \infty. \tag{B.17}$$

Then, we apply the selection vector again,

$$\begin{aligned} N^{-1}\|\Gamma_\ell^R\|^2 &= N^{-1}\|\Gamma^R a_\ell\|^2 = N^{-1}\|\Psi^R a_\ell - \Lambda^R a_\ell\|^2 = \left[N^{-1/2}\|\Psi^R a_\ell - \Lambda^R a_\ell\|\right]^2 \\ &\geq \left(N^{-1/2}\|\Psi^R a_\ell\| - N^{-1/2}\|\Lambda^R a_\ell\|\right)^2 \\ &= \left[(a_\ell' V_b a_\ell)^{1/2} - (a_\ell' V_a a_\ell)^{1/2}\right]^2 \\ &= \left[(\rho_\ell(\Sigma_\Psi \Sigma_{\overline{F}}))^{1/2} - (\rho_\ell(\Sigma_\Lambda \Sigma_F))^{1/2}\right]^2 + o(1). \end{aligned} \tag{B.18}$$

where the inequality follows from the triangle inequality, and the last equality holds since we can draw the similar conclusion from (B.4). Therefore, we have

$$N^{-1/2}\|\Gamma_\ell^R\| \geq \left|(\rho_\ell(\Sigma_\Psi \Sigma_{\overline{F}}))^{1/2} - (\rho_\ell(\Sigma_\Lambda \Sigma_F))^{1/2}\right| + o(1) \text{ for } \ell = 1, \cdots, r_a. \tag{B.19}$$

Recall Theorem 3.3(iii), we known that for $\ell = 1, \cdots, r_b$ and $\Gamma^0 \neq 0$,

$$N^{-1/2}\|\widehat{\Gamma}_\ell - \Gamma_\ell^R\| = O_p(C_{NT}^{-1}). \tag{B.20}$$

Then, based on (B.17), (B.19) and (B.20) and Assumption 1, we have for $\ell = 1, \cdots, r_b$,

$$\begin{aligned} \left|N^{-1/2}\|\widehat{\Gamma}_\ell\| - N^{-1/2}\|\Gamma_\ell^R\|\right| &\leq N^{-1/2}\|\widehat{\Gamma}_\ell - \Gamma_\ell^R\| = O_p(C_{NT}^{-1}), \\ -N^{-1/2}\|\widehat{\Gamma}_\ell - \Gamma_\ell^R\| &\leq N^{-1/2}\|\widehat{\Gamma}_\ell\| - N^{-1/2}\|\Gamma_\ell^R\|, \\ N^{-1/2}\|\Gamma_\ell^R\| - N^{-1/2}\|\widehat{\Gamma}_\ell - \Gamma_\ell^R\| &\leq N^{-1/2}\|\widehat{\Gamma}_\ell\|, \\ \left|(\rho_\ell(\Sigma_\Psi \Sigma_{\overline{F}}))^{1/2} - (\rho_\ell(\Sigma_\Lambda \Sigma_F))^{1/2}\right| + o(1) &\leq N^{-1/2}\|\widehat{\Gamma}_\ell\| + O_p(C_{NT}^{-1}). \end{aligned}$$

Therefore,

$$\lim_{N,T\to\infty} \Pr(\|\widehat{\Gamma}_\ell\| > 0) = 1 \text{ for } \ell = 1, \cdots, r_b. \tag{B.21}$$

Then, we consider the relation of two events $\{\|\widehat{\Gamma}_\ell\| > 0\} \subset \{\|\widehat{\Gamma}\| > 0\}$. Then we have

$$\Pr\big(\|\widehat{\Gamma}_\ell\| > 0\big) \leq \Pr\big(\|\widehat{\Gamma}\| > 0\big),$$

then

$$\lim_{N,T\to\infty} \Pr\big(\|\widehat{\Gamma}_\ell\| > 0\big) \leq \lim_{N,T\to\infty} \Pr\big(\|\widehat{\Gamma}\| > 0\big) \leq 1,$$

and then

$$1 \leq \lim_{N,T\to\infty} \Pr\big(\|\widehat{\Gamma}\| > 0\big) \leq 1,$$

this gives

$$\lim_{N,T\to\infty} \Pr\big(\|\widehat{\Gamma}\| > 0\big) = 1. \tag{B.22}$$

Moreover, since $\Pr(\|\widehat{\Gamma}\| > 0) + \Pr(\|\widehat{\Gamma}\| = 0) = 1$ and the two events $\{\|\widehat{\Gamma}\| > 0\}$ and $\{\|\widehat{\Gamma}\| = 0\}$ are disjoint, we apply the law of total probability, (B.21) and (B.22) to obtain

$$\begin{aligned}
\Pr\big(\mathcal{I}_{\{\|\widehat{\Gamma}_\ell\|>0\}} = 1\big) &= \Pr\Big(\{\mathcal{I}_{\{\|\widehat{\Gamma}_\ell\|>0\}} = 1\} \,\cap\, \{\|\widehat{\Gamma}\| > 0\}\Big) \\
&\quad + \Pr\Big(\{\mathcal{I}_{\{\|\widehat{\Gamma}_\ell\|>0\}} = 1\} \,\cap\, \{\|\widehat{\Gamma}\| = 0\}\Big), \\
\Pr\big(\mathcal{I}_{\{\|\widehat{\Gamma}_\ell\|>0\}} = 1\big) &= \Pr\Big(\{\|\widehat{\Gamma}_\ell\| > 0\} \,\cap\, \{\|\widehat{\Gamma}\| > 0\}\Big) \\
&\quad + \Pr\Big(\{\|\widehat{\Gamma}_\ell\| > 0\} \,\cap\, \{\|\widehat{\Gamma}\| = 0\}\Big), \\
\Pr\big(\mathcal{I}_{\{\|\widehat{\Gamma}_\ell\|>0\}} = 1\big) &= \Pr\Big(\{\|\widehat{\Gamma}_\ell\| > 0\}\Big) + \Pr\Big(\{\|\widehat{\Gamma}_\ell\| > 0\} \,\cap\, \{\|\widehat{\Gamma}\| = 0\}\Big),
\end{aligned}$$

$$\lim_{N,T\to\infty} \Pr\big(\mathcal{I}_{\{\|\widehat{\Gamma}_\ell\|>0\}} = 1\big) = \lim_{N,T\to\infty} \Pr\Big(\{\|\widehat{\Gamma}_\ell\| > 0\}\Big) + \lim_{N,T\to\infty} \Pr\Big(\{\|\widehat{\Gamma}_\ell\| > 0\} \cap \{\|\widehat{\Gamma}\| = 0\}\Big),$$

Therefore,

$$\lim_{N,T\to\infty} \Pr\big(\mathcal{I}_{\{\|\widehat{\Gamma}_\ell\|>0\}} = 1\big) = 1. \tag{B.23}$$

Similarly, we have $\big\{\mathcal{I}_{\{\|\widehat{\Gamma}_\ell\|>0\}} = 1\big\} \subset \big\{\mathcal{I}_{\{\|\widehat{\Gamma}\|>0\}} = 1\big\}$ as well as (B.23), which imply that

$$\begin{aligned}\Pr\big(\mathcal{I}_{\{\|\widehat{\Gamma}_\ell\|>0\}} = 1\big) &\le \Pr\big(\mathcal{I}_{\{\|\widehat{\Gamma}\|>0\}} = 1\big) \le 1\\ \lim_{N,T\to\infty} \Pr\big(\mathcal{I}_{\{\|\widehat{\Gamma}_\ell\|>0\}} = 1\big) &\le \lim_{N,T\to\infty} \Pr\big(\mathcal{I}_{\{\|\widehat{\Gamma}\|>0\}} = 1\big) \le 1\\ 1 &\le \lim_{N,T\to\infty} \Pr\big(\mathcal{I}_{\{\|\widehat{\Gamma}\|>0\}} = 1\big) \le 1.\end{aligned}$$

Therefore,

$$\lim_{N,T\to\infty} \Pr\big(\mathcal{I}_{\{\|\widehat{\Gamma}\|>0\}} = 1\big) = 1. \tag{B.24}$$

By (B.24) and the definition of $\widehat{\mathcal{B}}$ in (3.8), we conclude that

$$\lim_{N,T\to\infty} \Pr(\widehat{\mathcal{B}} = \mathcal{B}_0 = 1) = 1. \tag{B.25}$$

Hence, by taking into account for the break point and the instability, we have Theorem 3.3 proved by (B.12), (B.25) and (B.15).

Now, for the case where there is no change, i.e. $r_a = r_b$ and $\mathcal{B}_0 = 0$, Theorem 3.3(vi)

and (v) imply that

$$\lim_{N,T\to\infty} \Pr(\|\widehat{\Gamma}_\ell\| = 0) = 1 \text{ for } \ell = 1, \cdots, k, \tag{B.26}$$

By applying the same procedure from (B.23) and (B.24), we conclude that

$$\lim_{N,T\to\infty} \Pr\big(\mathcal{I}_{\{\|\widehat{\Gamma}\|>0\}} = 0\big) = 1. \tag{B.27}$$

Thus, by the definition of $\widehat{\mathcal{B}}$ in (3.8) and the fact that $\mathcal{B}_0 = 0$, we have

$$\lim_{N,T\to\infty} \Pr(\widehat{\mathcal{B}} = \mathcal{B}_0 = 0) = 1. \tag{B.28}$$

Then, the definition of $\widehat{r}_b$ in (3.10), the fact that $r_a = r_b$ and (B.12) imply that

$$\lim_{N,T\to\infty} \Pr(\widehat{r}_b = r_b) = 1. \tag{B.29}$$

Similarly, when there is no change, we have Theorem 3.4 proved based on (B.12), (B.28) and (B.29).

□

Appendix C

Supplemental Tables for the Great Recession Data Set

For the convenience of the reader, we report Table C.1 to C.4 below, which give the financial indicators of the Great Recession dataset analysed in Section 5.2. The reported tables are the same as given in Cheng et al. (2016, see Supplemental Appendix Tables S3-S5)[14].

Table C.1: List of Indicators - Part I

Name	Long Description
Cons: Dur	Real Personal Consumption Expenditures: Durable Goods
Cons: Svc	Real Personal Consumption Expenditures: Services
Cons: NonDur	Real Personal Consumption Expenditures: Non-durable Goods
Real InvtCh	Component for Change in Private Inventories, deflated by JCXFE
Real WageG	Component for Government GDP: Wage and Salary Disbursements by Industry, Government, deflated by JCXFE

Table C.2: List of Indicators - Part II

Name	Long Description
IP: DurGds materials	Industrial Production: Durable Materials
IP: NondurGds materials	Industrial Production: Nondurable Materials
IP: Auto	IP: Automotive products
IP: NonDurConsGoods	Industrial Production: Nondurable Consumer Goods
IP: BusEquip	Industrial Production: Business Equipment
IP: EnergyProds	IP: Consumer Energy Products
CapU Tot	Capacity Utilization: Total Industry
CapU Man	Capacity Utilization: Manufacturing (FRED past 1972)
Emp: DurGoods	All Employees: Durable Goods Manufacturing
Emp: Const	All Employees: Construction
Emp: Edu & Health	All Employees: Education & Health Services
Emp: Finance	All Employees: Financial Activities
Emp: Infor	All Employees: Information Services
Emp: Bus Serv	All Employees: Professional & Business Services
Emp: Leisure	All Employees: Leisure & Hospitality
Emp: OtherSvcs	All Employees: Other Services
Emp:Mining/NatRes	All Employees: Natural Resources & Mining
Emp: Trade&Trans	All Employees: Trade, Transportation & Utilities
Emp: Retail	All Employees: Retail Trade
Emp: Wholesal	All Employees: Wholesale Trade
Emp: Gov(Fed)	All Employees: Government: Federal
Emp: Gov (State)	All Employees: Government: State Government
Emp: Gov (Local)	All Employees: Government: Local Government
URate: Age16-19	Unemployment Rate - 16-19 yrs
URate: Age > 20 Men	Unemployment Rate - 20 yrs. & over, Men
URate: Age > 20 Women	Unemployment Rate - 20 yrs. & over, Women
U: Dur < 5wks	Number Unemployed for Less than 5 Weeks
U: Dur 5-14wks	Number Unemployed for 5-14 Weeks
U: Dur > 15-26wks	Civilians Unemployed for 15-26 Weeks
U: Dur > 27wks	Number Unemployed for 27 Weeks & over
U: Job Losers	Unemployment Level - Job Losers
U: LF Reentry	Unemployment Level - Reentrants to Labor Force
U: Job Leavers	Unemployment Level - Job Leavers
U: New Entrants	Unemployment Level - New Entrants
Emp: SlackWk	Employment Level - Part-Time for Economic Reasons, All Industries
AWH Man	Average Weekly Hours: Manufacturing
AWH Privat	Average Weekly Hours: Total Private Industrie
AWH Overtime	Average Weekly Hours: Overtime: Manufacturing

Table C.3: List of Indicators - Part III

HPermits	New Private Housing Units Authorized by Building Permit
Hstarts: MW	Housing Starts in Midwest Census Region
Hstarts: NE	Housing Starts in Northeast Census Region
Hstarts: S	Housing Starts in South Census Region
Hstarts: W	Housing Starts in West Census Region
Constr. Contracts	Construction contracts (mil. sq. ft.) (Copyright, McGraw-Hill)
IP: DurConsGoods	Industrial Production: Durable Consumer Goods
Ret. Sale	Sales of retail stores (mil. Chain 2000 $)
Orders (DurMfg)	Mfrs new orders durable goods industries (bil. chain 2000 $)
Orders (ConsumerGoods/Mat.)	Mfrs new orders, consumer goods and materials (mil. 1982 $)
UnfOrders (DurGds)	Mfrs unfilled orders durable goods indus. (bil. chain 2000 $)
Orders (NonDefCap)	Mfrs new orders, nondefense capital goods (mil. 1982 $)
VendPerf	Index of supplier deliveries vendor performance (pct.)
MT Invent	Manufacturing and trade inventories (bil. Chain 2005 $)
PCED-MotorVec	Motor vehicles and parts
PCED-DurHousehold	Furnishings and durable household equipment
PCED-Recreation	Recreational goods and vehicles
PCED-OthDurGds	Other durable goods
PCED-Food-Bev	Food and beverages purchased for off-premises consumption
PCED-Clothing	Clothing and footwear
PCED-Gas-Enrgy	Gasoline and other energy goods
PCED-OthNDurGds	Other nondurable goods
PCED-Housing-Utilities	Housing and utilities
PCED-HealthCare	Health care
PCED-TransSvg	Transportation services
PCED-RecServices	Recreation services
PCED-FoodServ-Acc.	Food services and accommodations
PCED-FIRE	Financial services and insurance
PCED-OtherServices	Other services
PPI: FinConsGds	Producer Price Index: Finished Consumer Goods
PPI: FinConsGds(Food)	Producer Price Index: Finished Consumer Foods
PPI: IndCom	Producer Price Index: Industrial Commodities
PPI: IntMat	Producer Price Index: Intermediate Materials: Supplies & Components
NAPM ComPrice	NAPM COMMODITY PRICES INDEX (PERCENT)
Real Price: NatGas	PPI: Natural Gas, deflated by PCEPILFE
Real Price: Oil	PPI: Crude Petroleum, deflated by PCEPILFE
FedFunds	Effective Federal Funds Rate

Table C.4: List of Indicators - Part IV

TB-3Mth	3-Month Treasury Bill: Secondary Market Rate
BAA-GS10	BAA-GS10 Spread
MRTG-GS10	Mortg-GS10 Spread
TB6m-TB3m	tb6m-tb3m
GS1-TB3m	GS1-Tb3m
GS10-TB3m	GS10-Tb3m
CP-TB Spread	CP-Tbill Spread: CP3FM-TB3MS
Ted-Spread	MED3-TB3MS (Version of TED Spread)
Real C&I Loan	Commercial and Industrial Loans at All Commercial BanksDefl by PCEPILFE
Real ConsLoans	Consumer (Individual) Loans at All Commercial Banks Outlier Code because of change in data in April 2010 see FRB H8 ReleasDefl by PCEPILFE
Real NonRevCredit	Total Nonrevolving Credit Owned and Securitized, OutstandingDefl by PCEPILFE
Real LoansRealEst	Real Estate Loans at All Commercial BanksDefl by PCEPILFE
Real RevolvCredit	Total Revolving Credit OutstandingDefl by PCEPILFE
S&P500	S&PS COMMON STOCK PRICE INDEX: COMPOSITE (1941-43=10)
DJIA	COMMON STOCK PRICES: DOW JONES INDUSTRIAL AVERAGE
VXO	VXO (Linked by N. Bloom) .. Average daily VIX from 2009
Ex rate: Major	FRB Nominal Major Currencies Dollar Index (Linked to EXRUS in 1973:1)
Ex rate: Switz	FOREIGN EXCHANGE RATE: SWITZERLAND (SWISS FRANC PER USD)
Ex rate: Japan	FOREIGN EXCHANGE RATE: JAPAN (YEN PER USD)
Ex rate: UK	FOREIGN EXCHANGE RATE: UNITED KINGDOM (CENTS PER POUND)
EX rate: Canada	FOREIGN EXCHANGE RATE: CANADA (CAD PER USD)

Bibliography

[1] Seung C. Ahn and Alex R. Horenstein. (2013). Eigenvalue Ratio Test for the Number of Factors. *Econometrica*, 81(3): 1203-1227.

[2] Alen Alexanderian. (2013). On Continuous Dependence of Roots of Polynomials on Coefficients. The University of Texas at Austin, USA.

[3] Dante Amengual and Mark W. Watson. (2007). Consistent Estimation of the Number of Dynamic Factors in a Large N and T Panel. *Journal of Business & Economic Statistics*, 25(1): 91-96.

[4] Jushan Bai. (1997). Estimation of A Change Point in Multiple Regression Models. *Review of Econometrics and Statistics*, 79(4): 551-563.

[5] Jushan Bai and Serena Ng. (2002). Determining the Number of Factors in Approximate Factor Models. *Econometrica*, 70(1): 191-221.

[6] Brandon J. Bates, Mikkel Plagborg-Møller, James H. Stock and Mark W. Watson. (2013). Consistent Factor Estimation in Dynamic Factor Models with Structural Instability. *Journal of Econometrics*, 177(2): 289-304.

[7] Ben S. Bernanke, Jean Boivin and Piotr Eliasz. (2005). Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach. *The Quarterly Journal of Economics*, 120(1): 387-422.

[8] Peter J. Bickel and Kjell A. Doksum. (2001). Mathematical Statistics, Basic Ideas and Selected Topics. Volume 1, Second Edition. New Jersey: Prentice Hall.

[9] Carl de Boor (2002). Applied Linear Algebra. UW-Madison Department of Mathematics Publications.

[10] Jörg Breitung and Sandra Eickmeier. (2011). Testing for Structural Breaks in Dynamic Factor Models. *Journal of Econometrics*, 163(1): 71-84.

[11] Peter Bühlmann and Sara van de Geer. (2011). Statistics for High-Dimensional Data: Methods, Theory and Applications. New York: Springer.

[12] George Casella and Roger L. Berger. (2002). Statistical Inference. Second Edition. Duxbury Press.

[13] Liang Chen, Juan J. Dolado and Jesús Gonzalo. (2014). Detecting Big Structural Breaks in Large Factor Models. *Journal of Econometrics*, 180(1): 30-48.

[14] Xu Cheng, Zhipeng Liao and Frank Schorfheide. (2016). Shrinkage Estimation of High-Dimensional Factor Models with Structural Instabilities. *Review of Economic Studies*, 83(4): 1511-1543. Oxford University Press.

[15] Stephen G. Donald. (1997). Inference Concerning the Number of Factors in a Multivariate Nonparametric Relationship. *Econometrica*, 65(1): 103-132.

[16] Bradley Efron and Robert J. Tibshirani. (1993) An Introduction to the Bootstrap. Chapman & Hall/CRC

[17] Jianqing Fan, Jinchi Lv and Lei Qi. (2011). Sparse High Dimensional Models in Economics. *Annual Review of Economics*, 2011 Sep(3): 291-317.

[18] Wayne A. Fuller. (1976). Introduction to Statistical Time Series. John Wiley.

[19] David B. Grusky, Bruce Western and Christopher Wimer. (2011). The Great Recession. New York: Russell Sage Foundation.

[20] Domenico Giannone, Lucrezia Reichlin and David Small. (2008). Nowcasting: The Real-Time Informational Content of Macroeconomic Data. *Journal of Monetary Economics*, 55(2008):665-676.

[21] Arthur Lewbel. (1991). The Rank of Demand Systems: Theory and Nonparametric Estimation. *Econometrica*, 59(3): 711-730.

[22] Xu Han and Atsushi Inoue. (2014). Tests for Parameter Instability in Dynamic Factor Models. *Econometric Theory*, 31(5): 1117-1152.

[23] Sankaran Mahadevan and Animesh Dey. (1997). Adaptive Monte Carlo Simulation for Time-Variant Reliability Analysis of Brittle Structures. *AIAA Journal*, 35(2): 321-326.

[24] Christopher Z. Mooney. (1997). Monte Carlo Simulation. Sage University Paper series on Quantitative Applications in the Social Science, 07-116. Thousand Oaks, CA: Sage.

[25] Sévérien Nkurunziza, Marwan Al-Momani, and Eric Y. Y. Lin. (2016). Shrinkage and LASSO Strategies in High Dimensional Heteroscedastic Models. *Communications in Statistics − Theory and Methods*, 45(15): 4454-4470.

[26] Alexei Onatski. (2010). Determining the Number of Factors from Empirical Distribution of Eigenvalues. *Review of Economics and Statistics*, 92(4): 1004-1016.

[27] Michael Roberts. (2009). The Great Recession - Profit Cycles, Economic Crisis, A Marxist View. Lulu Press.

[28] Peter H. Schönemann. (1966). A Generalized Solution of the Orthogonal Procrustes Problem. *Psychometrika*, 31(1): 1-10.

[29] Christopher A. Sims, James H. Stock and Mark W. Watson. (1990). Inference in Linear Time Series Models with Some Unit Roots. *Econometrica*, 58(1): 113-144.

[30] Robert Stine. (1989). An Introduction to Bootstrap Methods - Examples and Ideas. *Sociological Methods and Research*, 18(2-3): 243-291.

[31] James H. Stock and Mark W. Watson. (2012). Disentangling the Channels of the 2007-09 Recession. *Brookings Papers on Economic Activity*, 81-156.

[32] Robert L. Strawderman. (1994). A Note on Necessary and Sufficient Conditions for Proving That a Random Symmetric Matrix Converges to A Given Limit. *Statistics & Probability Letters*, 21: 367-370.

[33] Xianchao Xie, S. C. Kou and Lawrence Brown. (2016). Optimal Shrinkage Estimation of Mean Parameters in Family of Distributions with Quadratic Variance. *The Annals of Statistics*, 44(2): 564-597.

[34] Ming Yuan and Yi Lin. (2006). Model Selection and Estimation in Regression with Grouped Variables. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 68(1): 49-67.

Vita Auctoris

Miss. Wen (Wendy) Gao was born in 1994, Dalian, Liaoning, China. She graduated from Dalian No.1 High School in June 2012. From there she went to University of California at Davis, where she obtained a Bachelor of Science degree in Statistics. She is currently a candidate for the Master of Science degree in Statistics at the University of Windsor. She hopes to graduate in September 2018.